

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:24:19 ; Search time 16.6136 Seconds
(without alignments)

3048.258 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MITYTSQVANRLGSPSRLL.....QLLPSSVBPPLWAPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pap.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	361	15.4	10	US-09-768-826-54
2	239	10.2	251	Sequence 54, Appl
3	238	10.1	99	Sequence 35, Appl
4	151	6.4	118	Sequence 429, Appl
5	137	5.8	60	US-09-764-877-2006
6	135	5.7	71	Sequence 2006, Ap
7	134	5.7	65	Sequence 1622, Ap
8	130	5.5	40	Sequence 243, Ap
9	130	5.5	76	Sequence 6346, Ap
10	130	5.5	76	Sequence 6318, Ap
11	130	5.5	76	Sequence 575, Ap
12	130	5.5	76	Sequence 888, Ap
13	130	5.5	76	Sequence 575, Ap
14	130	5.5	76	Sequence 888, Ap
15	130	5.5	76	Sequence 575, Ap
16	130	5.5	76	Sequence 888, Ap
17	130	5.5	76	Sequence 575, Ap
18	130	5.5	76	Sequence 888, Ap
19	130	5.5	76	Sequence 575, Ap

20	130	5.5	76	10	US-09-822-827-888	Sequence 888, App
21	129.5	5.5	45	9	US-10-106-698-4849	Sequence 4849, Ap
22	129	5.5	72	9	US-09-925-299-1027	Sequence 1027, Ap
23	129	5.5	72	10	US-09-925-299-1027	Sequence 1027, Ap
24	125	5.3	47	9	US-10-106-698-4727	Sequence 4727, Ap
25	124.5	5.3	66	9	US-09-764-891-4710	Sequence 4710, Ap
26	121	5.1	60	9	US-09-764-872-376	Sequence 376, App
27	120	5.1	104	9	US-10-073-961-260	Sequence 260, App
28	120	5.1	104	10	US-09-764-887-260	Sequence 260, App
29	120	5.1	123	9	US-10-092-154-509	Sequence 509, App
30	120	5.1	123	10	US-09-764-847-509	Sequence 509, App
31	119	5.1	68	10	US-09-925-300-1301	Sequence 1301, App
32	118.5	5.0	74	9	US-10-001-835-210	Sequence 210, App
33	117.5	5.0	75	9	US-10-001-835-182	Sequence 182, App
34	117.5	5.0	75	9	US-10-001-835-195	Sequence 195, App
35	117	5.0	72	9	US-10-073-961-301	Sequence 301, App
36	117	5.0	72	10	US-09-764-887-301	Sequence 301, App
37	116	4.9	40	9	US-10-106-698-5126	Sequence 5126, Ap
38	114.5	4.9	75	9	US-10-001-835-225	Sequence 225, App
39	114.5	4.9	287	10	US-09-799-983-4	Sequence 4, Appli
40	114.5	4.9	344	10	US-09-799-983-6	Sequence 6, Appli
41	114.5	4.9	344	10	US-09-799-983-2	Sequence 2, Appli
42	113	4.8	94	9	US-10-091-504-998	Sequence 998, App
43	113	4.8	94	10	US-09-764-869-998	Sequence 998, App
44	111.5	4.7	75	9	US-10-001-835-223	Sequence 223, App
45	111	4.7	52	9	US-09-983-802-537	Sequence 537, App

ALIGNMENTS

RESULT 1
US-09-768-826-54
; Sequence 54, Application US/09768826
; Patent No. US2002001296A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512PI
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-54

Query Match	15.4%	Score 361;	DB 10;	Length 314;
Best Local Similarity	43.6%	Pred. No. 4.2e-25;		
Matches	88;	Conservative 14;	Mismatches 78;	Indels 22; Gaps 4;
QY	172	EXLSLPHNMFWVPWFWEANLSMKWLGGRIRDPILLOSLINEMNTLTTCGHLIYADWIS	231	
Db	13	ESLSKSDFNKYWPCVWFNTLAAQARDGRIRDDIALCLLLELNKYRACSKMLFHYDWIS	72	
QY	232	IFLVYTVQVTVAVYSGFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLVVPVF	276	
Db	73	IFLVYTVQVTVAVYSGFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLVVPVF	276	
QY	277	TFLOPFFVGMKVLGSLRALLGWRHGQGHGOQLLETRMOCQERKVSrv-----ESSQ	329	
Db	133	TLLOPFFVGMKVLGSLRALLGWRHGQGHGOQLLETRMOCQERKVSrv-----ESSQ	329	
QY	330	AWRTRPVIPATREAEAGESLEP	351	
Db	193	YDEDPQPPYTVATAESLRP	214	

```
RESULT 2
US-09-768-826-35
; Sequence 35, Application US/09768826
; Patent No. US2002012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PFS12P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-35

Query Match          10.2%; Score 239; DB 10; Length 251;
Best Local Similarity 42.0%; Pred. No. 5.3e-14;
Matches 63; Conservative 9; Mismatches 56; Indels 22; Gaps 4;

QY 224 LYAYDWISIPLVYTVVTVAVISFFLTCLVGRQFLNP-----AKAY-----PGHE----- 268
Db 2 LFHYDWISIPLVYTVVTVAVISFFLTCLVGRQFLNP-----AKAY-----PGHE----- 268

QY 269 LBLVVPVFTLQFFVTVGVLKVLGRALLGVRHGRGQGLLETRMQCQERKVSRY--- 325
Db 62 PMYVPLTLLQFFVTVGVLKVLGRALLGVRHGRGQGLLETRMQCQERKVSRY--- 325

QY 326 -----ESSQAWRTPTVPATREAGSLESP 351
Db 122 LPAEKDQWDEDDQPPPTVATAESLRP 151

RESULT 3
US-09-764-872-429
; Sequence 429, Application US/09764872
; Publication No. US2003050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 429
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-429

Query Match          10.1%; Score 238; DB 9; Length 99;
Best Local Similarity 55.4%; Pred. No. 2e-14;
Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFSLKLCWRGSIYKLYGBFLFLCLCYIIRFYRLALTEEQOL 60
Db 13 MTSVYTLKVAERFGFGLLRWRGSIYKLYKEFLFLGALYAVLSITYRLLLTQEQR 72

QY 61 MPEKLTLCDSYIQLPISFVLG 83
Db 73 VYAQVARYCNRSADLPISFVLG 95

RESULT 4
US-09-764-877-2006
; Sequence 2006, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2006
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-2006

Query Match          6.4%; Score 151; DB 10; Length 118;
Best Local Similarity 37.3%; Pred. No. 2.5e-06;
Matches 38; Conservative 4; Mismatches 26; Indels 34; Gaps 4;

QY 329 QAWWRTPTVPATREAGSLESPGRRRLWQSSSTPLERMMMLRPTGLSTGICPCW 388
Db 12 QAWWMAVVPVATREAGSELLEPGRRRL-----HSEP----- 43

QY 389 LWMRCTRTCLGWSRTCTGSPSHSPPTQLLPSSVPEPLWAP 430
Db 44 ---RLHHCPTGWATECDVSVKQK-OKQLPPPPQTPP---WSP 79

RESULT 5
US-09-764-877-1622
; Sequence 1622, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1622
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1622

Query Match          5.8%; Score 137; DB 10; Length 60;
Best Local Similarity 68.6%; Pred. No. 2e-05;
Matches 24; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 322 VSRVSSQAWWRTPTVPATREAGSLESPGRRRL 356
Db 5 IKNIKISQAWWRMPLVPATREAGSLESGRQL 39

RESULT 6
US-10-011-585A-243
; Sequence 243, Application US/10011585A
; Publication No. US20030039986A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
```

```
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0261
; CURRENT APPLICATION NUMBER: US/10/011.585A
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/245,740
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 245
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 243
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-011-585A-243

Query Match      5.7%; Score 135; DB 9; Length 71;
Best Local Similarity 44.1%; Pred. No. 3.8e-05;
Matches 30; Conservative 7; Mismatches 17; Indels 14; Gaps 2;

QY 286 GMLKVLGRALLGWRHGRGHGQQLLETRMQCQKVRVSSQAWWRTVPVATREAEA 345
Db 18 GMIAGQEFNAGSGHG-----ETPSLLKIQKISR-----TWQAPVIPATQEA 63

QY 346 GESLEPR 353
Db 64 GESLEPR 71

RESULT 7
US-10-106-698-6346
; Sequence 6346, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6346
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6346

Query Match      5.7%; Score 134; DB 9; Length 65;
Best Local Similarity 54.5%; Pred. No. 4.2e-05;
Matches 30; Conservative 4; Mismatches 11; Indels 10; Gaps 2;

QY 302 GQRHGQQLLETRMQCQKVRVSSQAWWRTVPVATREAGESLEPGRRRL 356
Db 20 GQDGETPSLLKI-----QKKISR-----AWWHVPVIPATWETEAGELLEPGRRRL 64

RESULT 8
US-10-106-698-6318
; Sequence 6318, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
```

```
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6318
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6318

Query Match      5.5%; Score 130; DB 9; Length 40;
Best Local Similarity 69.4%; Pred. No. 5.3e-05;
Matches 25; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 321 KVSRRVSSQAWWRTVPVATREAGESLEPGRRRL 356
Db 6 KIQKI--SWAWWRTVPVATLEAAGESLAPRRRL 39

RESULT 9
US-10-012-896-575
; Sequence 575, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Panger, Gary R.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-575

Query Match      5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 318 QKRVSRVSSQAWWRTVPVATREAGESLEPGRRRL 355
Db 3 KSRFTKNTKIQAWWRAPVTPCTREAGESLEPGRLR 40

RESULT 10
US-10-012-896-888
```

Sequence 888, Application US/10012896
Publication No. US20020183251A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqiu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
APPLICANT: Wantanabe, Yoshihiro
APPLICANT: Meagher, Madeleine Joy

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.427C27

CURRENT APPLICATION NUMBER: US/10/012,896

CURRENT FILING DATE: 2001-12-10

NUMBER OF SEQ ID NOS: 1011

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 888

LENGTH: 76

TYPE: PRT

ORGANISM: Homo sapiens

US-10-012-896-888

Query Match 5.5%; Score 130; DB 9; Length 76;

Best Local Similarity 63.2%; Pred. No. 0.00012;

Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRR 355

Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 11

US-09-895-793-575

Sequence 575, Application US/09895793

Publication No. US20020192763A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqiu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.534C2

CURRENT APPLICATION NUMBER: US/09/895,793

CURRENT FILING DATE: 2001-06-29

NUMBER OF SEQ ID NOS: 982

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 888

LENGTH: 76

TYPE: PRT

ORGANISM: Homo sapiens

US-09-895-793-888

APPLICANT: Vinals de Bassols, Carlota

APPLICANT: Foy, Teresa

APPLICANT: Fanger, Gary R.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.534C2

CURRENT APPLICATION NUMBER: US/09/895,793

CURRENT FILING DATE: 2001-06-29

NUMBER OF SEQ ID NOS: 982

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 575

LENGTH: 76

TYPE: PRT

ORGANISM: Homo sapiens

US-09-895-793-575

Query Match 5.5%; Score 130; DB 9; Length 76;

Best Local Similarity 63.2%; Pred. No. 0.00012;

Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRR 355

Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 12

US-09-895-793-888

Sequence 888, Application US/09895793

Publication No. US20020192763A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Harlocker, Susan L.

APPLICANT: Jiang, Yuqiu

APPLICANT: Kalos, Michael D.

APPLICANT: Retter, Marc W.

APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Darrick

APPLICANT: Li, Samuel X.

APPLICANT: Wang, Aijun

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

APPLICANT: Hural, John

APPLICANT: McNeill, Patricia D.

APPLICANT: Houghton, Raymond L.

APPLICANT: Vinals de Bassols, Carlota

APPLICANT: Foy, Teresa

APPLICANT: Fanger, Gary R.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.534C2

CURRENT APPLICATION NUMBER: US/09/895,793

CURRENT FILING DATE: 2001-06-29

NUMBER OF SEQ ID NOS: 982

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 888

LENGTH: 76

TYPE: PRT

ORGANISM: Homo sapiens

US-09-895-793-888

Query Match

Best Local Similarity 5.5%; Score 130; DB 9; Length 76;

Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 318 QERKVSRYVSSQAWRTVPVTPATREAGGESLEPGRR 355

Db 3 KSRFTKNTKITQAWWRAPVPGTREAEGGESLEPGRLR 40


```
RESULT 13
US-09-895-814-575
; Sequence 575, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-575

Query Match      5.5%; Score 130; DB 9; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY      318 QERKVSRESSQAWRTVPVTPATREAEAGESLEPGRRR 355
Db      3 KSRFTKTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

RESULT 14
US-09-759-143-575
; Sequence 575, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-575

Query Match      5.5%; Score 130; DB 10; Length 76;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY      318 QERKVSRESSQAWRTVPVTPATREAEAGESLEPGRRR 355
Db      3 KSRFTKTKITQAWWRAPVPGTREAEGGESLEPGRLR 40

Search completed: July 10, 2003, 12:28:57
Job time : 16.6136 secs
```



```
RESULT 5
T21644
hypothetical protein F32B6.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21644
R:Basham, V.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19453
A:Accession: T21644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <MIL>
A:Cross-references: EMBL:Z81074; PIDN:CA803043.1; GSPDB:GN00022; CESP:F32B6.9
A:Experimental source: clone F32B6
C:Genetics:
A:Gene: CESP:F32B6.9
A:Map position: 4
A:Introns: 22/2; 61/1; 141/2; 176/3; 200/3; 290/3; 332/1

Query Match      24.4%; Score 573.5; DB 2; Length 413;
Best Local Similarity 39.9%; Pred. No. 2.3e-42;
Matches 120; Conservative 53; Mismatches 109; Indels 19; Gaps 5;

QY 1 MTITYTSQVANARLGFSRLLLCWRGSIYKLYGFIPLCYIIRFIYRLAL-----T 55
DQ 1 MTISYS-----GNVIRILLRWKGSIMRTAKKELLILYLYSVRVFYLKGDIDDD 52
QY 56 EQOOL---MFEKLTLYCDSYIQLIPISFVLGVFTLVVTRWNNQYENLPWDRMLSLVS 111
DQ 53 EDRLLKVRMFETFCQCDSYTFLIPLTLFLGFFVSVVAVRWVRQFETLYWDPEDILSVLC 112
QY 112 GFVEGKDEOGLRLRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQL 171
DQ 113 TVLHODEKSKRRHTIARYNLNANLAWRDISKTRLRPSVHSLIESGLLTKETKQIL 172
QY 172 EKLSLPH-NMFVWPVWFANLKMKGIRDPILLQSLNEMNTLRTOCGHLYAYDWI 230
DQ 173 EAMHAENESSRWITPLHWIQLMRQVEEHKPTASLNFQVGBELRIFRQSLRKLSDYDW 232
QY 231 SPLVYTVVTVAVYFFLTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLOFFYVGLKV 290
DQ 233 CVPLVYTVAAALATYSFFPFTLFGROFLP-DIETGKELDLVVPVFTFIVQFLFVGNFKV 291
QY 291 G 291
DQ 292 G 292

RESULT 6
T27630
hypothetical protein ZC518.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
C:Accession: T27630
R:Thomas, K.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z20396
A:Accession: T27630
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-499 <MIL>
A:Cross-references: EMBL:Z68753; PIDN:CAA92989.1; GSPDB:GN00022; CESP:ZC518.1
A:Experimental source: clone ZC518
C:Genetics:
A:Gene: CESP:ZC518.1
A:Map position: 4
A:Introns: 30/3; 60/2; 106/2; 239/3; 267/3; 315/3; 329/3; 380/3; 445/2; 466/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match      24.3%; Score 571; DB 2; Length 499;
Best Local Similarity 32.4%; Pred. No. 4.7e-42;
```

```
Matches 124; Conservative 70; Mismatches 145; Indels 44; Gaps 6;

QY 1 MTITYTSQVANARLGFSRLLLCWRGSIYKLYGFIPLCYIIRFIYRLALTEEQOL 60
DQ 1 MTISYTLDVQTNLQSPFSLLRWGSVWKAQVQLAVTAVFLISCIYRMLSPSQOD 60
QY 61 MFEKLTLYCDSYIQL-IPISFVLGVFTLVVTRWNNQYENLPWDRMLSLVSGFVEGKDE 119
DQ 61 VFEQLIRYFDNKLDAIPLTLFLGFFSVFVAVRWGSLNGIGWIDDAALLPATYIRGADE 120
QY 120 QGRLRLRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKOLEKLSLPHN 179
DQ 121 ETRVIRNLRYVLVSQALVLRDISMQVRKRPFTMDLAAASGLMTHSEMDILDHDKDPSYS 180
QY 180 MFWVWPVWFANLKMKGIRDPILLQSLNEMNTLRTOCGHLYAYDWISIPLVYTVQV 239
DQ 181 RYWTSIQWLSNLVYECQKGVDSYLYLNNKIVDIBGKFRHGLASLLKLYDVPVFLVYQV 240
QY 240 TVVAVYFFLTCLVGRQFL-----NPAKAYPGHGLDLVVPVFTFLOFFYVGLKV----- 290
DQ 241 IFLAVRIYFIMCLIGRQFIVTGNPS-----GIDLWLPITTMVQFLVYMGWKMVABALL 294
QY 291 -----GLSRALLGWRHQGHGQQLLETRQCQKQKRVSRVRESS 328
DQ 295 NPLGEDDDDECNYIIDKNLITGLSVITWTKHDDTGY--SMVEEHM-----AKTPAQKKD 348
QY 329 QAWRTPVIPATREAEAGESLEP 351
DQ 349 EFWGIDKIAPLYSMESAESRVHP 371

RESULT 7
T19565
hypothetical protein C29F4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T19565
R:Kershaw, J.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19143
A:Accession: T19565
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-584 <MIL>
A:Cross-references: EMBL:Z68335; PIDN:CAA92730.1; GSPDB:GN00022; CESP:C29F4.2
A:Experimental source: clone C29F4
C:Genetics:
A:Gene: CESP:C29F4.2
A:Map position: 4
A:Introns: 36/2; 78/2; 104/2; 149/2; 174/2; 221/2; 256/3; 333/3; 412/2; 484/2; 526/1;
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match      23.8%; Score 559; DB 2; Length 584;
Best Local Similarity 36.9%; Pred. No. 6.3e-41;
Matches 101; Conservative 64; Mismatches 109; Indels 0; Gaps 0;

QY 19 RLILCWRGSIYKLYGFIPLCYIIRFIYRLALTEEQOLMFEKLTLYCDSYIQLPI 78
DQ 63 KLIFKWKSLWKAIYLDLIVWCFCYAFISVIVYALDRSQODTFERFMQFCNRRLDYPI 122
QY 79 SFVLGVFTLVVTRWNNQYENLPWDRMLSLVSGFVEGKDEOGLRLRLIRYANLGNVL 138
DQ 123 NFMVGFVTVINRWMTQANLGMIDNIALFTSYMLSGNDRGRILRSIVRCVMSQTM 182
QY 139 ILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKOLEKLSLPHNMFVWPVWFANLKMKAWL 198
DQ 183 VFRDIHIGVRKRPPTLETVAAGITSTSELKCYNEVESRYAKYWLGNFMTFNLLNEARRE 242
QY 199 GRIRDPILLQSLNEMNTLRTOCGHLYAYDWISIPLVYTVQVTVAVYFFLTCLVGRQFL 258
DQ 243 GRIESAYTONAIAEIRTFRSGLSLITWYDWPILMYPOLVFMVPMATHCYCYLVCLVSRQV 302
QY 259 NPAKAYPGHGLDLVVPVFTFLOFFYVGLKVGL 292
```


T18781
 hypothetical protein B0564.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T18781
 R:Lightning, J.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19021
 A:Accession: T18781
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-450 <MIL>
 A:Cross-references: EMBL:Z73422; PIDN:CAA97765.1; GSPDB:GN000022; CESP:B0564.3
 A:Experimental source: clone B0564
 C:Genetics:
 A:Gene: CESP:B0564.3
 A:Map position: 4
 A:Introns: 60/2; 105/2; 320/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 22.5%; Score 529; DB 2; Length 450;
 Best Local Similarity 37.3%; Pred. No. 2e-38;
 Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;

QY 1 MTITTSQVANRLGSPFLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQQL 60
 DB 1 MTINHKETIMTSPWTFLLPKWGSIKWAVYMETIIFLCYIIISVIYKTAMGESSQR 60

QY 61 MFEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWDLMSLVSGFVEGKDEQ 120
 DB 61 TRESLVRYFDKLSYIPLFVGLGFTTVTRWTKLYQTIGFDNVLGMANCYIRGATEK 120

QY 121 GRLLRRLTRYANLGNVLILRSVSTAVYKRPSPQAHLVQAGFMTPAEHKOLEKLSLPHN- 179
 DB 121 ARIYRNIMRYCELVLQILVFRDMSMRTRRPPTMTVVAAGFMNKHELELYNSYDTKYN 180

QY 180 ----MFVWPVWVFWPANLSMKAWLGGRIKDPILLOSLNEMNTLRTOCHLYAYDWISPLV 235
 DB 181 KLGTKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIEWVCNDWVPLPM 240

QY 236 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLVKV 290
 DB 241 YPOLVCLAVNLVFLVSIIRQ-LVIEKHKNVDEVDVYFPWTFLOFIFYNGMLKV 294

RESULT 12
 T18782
 hypothetical protein B0564.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T18782
 R:Lightning, J.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19021
 A:Accession: T18782
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-523 <MIL>
 A:Cross-references: EMBL:Z73422; PIDN:CAA97766.1; GSPDB:GN000022; CESP:B0564.4
 A:Experimental source: clone B0564
 C:Genetics:
 A:Gene: CESP:B0564.4
 A:Map position: 4
 A:Introns: 60/2; 105/2; 320/3; 405/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 22.5%; Score 528; DB 2; Length 523;
 Best Local Similarity 37.6%; Pred. No. 2.9e-38;
 Matches 111; Conservative 52; Mismatches 126; Indels 6; Gaps 2;

QY 1 MTITTSQVANRLGSPFLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQQL 60
 DB 1 MTINHKETIKTSHWTKFFVLLFRWGSIKWAIYMETIIFLCYIIISVIYKTAMGESSQR 60

QY 61 MFEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWDLMSLVSGFVEGKDEQ 120
 DB 61 TRESVIRYCDKRLSFIPLFVGLGFTTVTRWTKLARTVGFIDVCLANLYVGTSEK 120

QY 121 GRLLRRLTRYANLGNVLILRSVSTAVYKRPSPQAHLVQAGFMTPAEHKOLEKLSLPHN- 179
 DB 121 AIIYRNIRIARYCALTOQLVFRDMSMRTRRPPTMTVVAAGFMNKDELDLYNSYTKYNS 180

QY 180 ----MFVWPVWVFWPANLSMKAWLGGRIKDPILLOSLNEMNTLRTOCHLYAYDWISPLV 235
 DB 181 RLGGKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIEWVCNDWVPLPM 240

QY 236 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLVKV 290
 DB 241 YPOLVCLAVNLVFLVSIIRQ-LVIEKHKNVDEVDVYFPWTFLOFIFYNGMLKV 294

RESULT 13
 S40708
 hypothetical protein C07A9.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Aug-2000
 C:Accession: S40708
 R:Smith, M.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: S40701
 A:Accession: S40708
 A:Molecule type: DNA
 A:Residues: 1-459 <SMI>
 A:Cross-references: EMBL:Z29094; NID:g436440; PID:g436448
 C:Genetics:
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.8%; Score 512; DB 2; Length 459;
 Best Local Similarity 31.9%; Pred. No. 6.3e-37;
 Matches 95; Conservative 71; Mismatches 126; Indels 6; Gaps 1;

QY 1 MTITTSQVANRLGSPFLLCWGSIYKLLYGEFLIFLLCYIIIRFYRLALTEEQ- 59
 DB 44 LSYNYVDIATSKSLMIVRMIFKRWGSLQVAKELIWIICAYLSVSVIRFALTRSQKE 103

QY 60 ----LMFEKLTLYCDSYIQLIPISFVLGFTVTVTRWNOYENLPWDLMSLVSGFV 114
 DB 104 QNKEIIIFERFGCYCDARMGYLPLNFVGLGFTTVTRWTKLYTSLGNIDNIALFVSAYV 163

QY 115 EGKDEQRLRLTRYANLGNVLILRSVSTAVYKRPSPQAHLVQAGFMTPAEHKOLEK 174
 DB 164 RGTDDRARQIRRNIRIARYCVISQCLVFRDIHVGVRRRPPTLEAVAQAGIMLPHELEKFSI 223

QY 175 SLPHNMFVWPVFWPANLSMKAWLGGRIKDPILLOSLNEMNTLRTOCHLYAYDWISPL 234
 DB 224 KRIYQKIVWVFWNALLELNVAKTEKIDGNARNAAQEIISFRSALTTVSMYDWDVPIPL 283

QY 235 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLVKVCL 292
 DB 284 MYPQLVNMVAVHYFFLCITRQFFISADAHNKTEVDLYTFPMTIIEIFIFNGMLKVAM 341

RESULT 14
 B88710
 protein C43G2.4 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Mar-2002
 C:Accession: B88710
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.el
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
 A:Accession: B88710

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:chr_IV; PIDN:AB091111.1; PID:g1572760; GSPDB:GN00022; CESP:C43G2.
C:Genetics:
A:Gene: C43G2.4
A:Map position: 4
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.3%; Score 500.5; DB 2; Length 420;
Best Local Similarity 29.2%; Pred. No. 5.7e-36;
Matches 131; Conservative 71; Mismatches 125; Indels 121; Gaps 15;

QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLYGELFLICLVYIRF----- 48
DB 1 MTISYS-----GNFFRLLLAWKGSINWSVRELFLFLFYFRFSPAFHFNYTDPT 52
QY 49 -----IYRLAITEQOOLMFEKLTLYCDSYIOLIPISFVLGPFYVTLVTRWNYENLP 101
DB 53 DSKGYRKIFKVMCNFHE-----YTKMPLTLFLGFGVSNVSRWVRQFETLR 100
QY 102 WPDRLMSLVSGFVEGDEQGLRRLTLIRYANLGNVLLRSVSTAVYKRPSPSAQHLVQAG 161
DB 101 WEDFLSILCLLPK--ESRPARHQIARYLINTSALAWRDVSTKRLRFPESURNIDAG 158
QY 162 FMTPAEHKQLEKLSLPHNMFV--PWVWFANL-SMKAWLGGRIKDPD-LIQSL----- 211
DB 159 LLTEYEKQLDINVSQAVILMLPVVTFKLCETSPSGIRWLTLPHWVQQLIDAEITA 218
QY 212 -----NEMNTLRTQCGHLYAYDMISPLVYQVTVVAVYSFFLTCLVGRQFLN 259
DB 219 GRGSVNVSVATNELKAYRISFRLLYCHDWCVPLVYTVQALATYSYFFCLFGQDLN 278
QY 260 PAKAYPGHELDLVVPVTFLOFFFYVGLKVG-----LSRALLG-- 298
DB 279 HDQFY---SLDAFFPLTVVQFLFFVGNFKVGQDLMRPFGLDDDDFELSILDRNIVTSF 335
QY 299 -----WRHQRGHGQQLLETRMQCQERKVSRYVSSQAWWRTVPVIPA 339
DB 336 TIVDSIQDDDDPPKFEEDVFWKH---HNEQ-----QQHQSMFLPVPTSLKNGR---IDL 384
QY 340 TREAEAGESLEPGRRLLWQSSSTPLE 367
DB 385 SRNAHK---HPPKLTLYLEMKNQDPEE 408

RESULT 15
S42371
hypothetical protein T20G5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S42371
R:Smith, A.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42368
A:Accession: S42371
A:Molecule type: DNA
A:Residues: 1-405 <SMI>
A:Cross-references: EMBL:Z30423; NID:G458479; PID:G458483
C:Genetics:
A:Introns: 23/2; 105/2; 267/3; 315/3; 366/2; 396/2; 404/1
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.0%; Score 494.5; DB 2; Length 405;
Best Local Similarity 30.2%; Pred. No. 1.8e-35;
Matches 109; Conservative 74; Mismatches 155; Indels 23; Gaps 6;

QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLYGELFLICLVYIRFIRYRLALTEEQOL 60
DB 1 MTYSVNSQSVATSRPWTFFLALIFRWGVSWSAIQYVNLGLYFLVSAIYRFLSAYQQQ 60
QY 61 MFEKLTLYCDSYIQLIPISFVLGPFYVTLVTRWNYENLPWPDRLMSLVSGFVEGKDEQ 120

Db 61 IFVRLVDVYNSRMSYVPLDMMLGFFFIAGVLRFRFWLYDIIGFDINTACSTATVIRGDSR 120
QY 121 GRLLRRTLIRYANLGNVLLRSVSTAVYKRPSPSAOHLVQAGFMTPAEHKQLEKLSLPHNM 180
DB 121 AKQYRNNIIRYCELTQVLI FRDLSMKARKRFPFTLDTVAAAAGFPMPEKAFDLIQYNYNK 180
QY 181 FVWPVWVFANLSMKAWLGGRIKDPDILQLLNMNTLRTQCGHLYAYDMISPLVYTVQV 240
DB 181 YELPFNWAVALYTNARKEGLIEGDIYTVVISDIKKFRTGLAWVCNVDWVPLEIITYPTIV 240
QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFFYVGLKVSRLALLG-W 299
DB 241 CLAVHMYFFVGILARQYKVGSIIDP-DMIDLVPFPMTSIQVFYMGMLKVG--EGLNPNW 297
QY 300 RHQQRCHGQQLLETRMQCQER-----KVSRYVSSQAWWRTVPVIPATREAEAG 346
DB 298 -----GEDPDDPETNNMLIDRNLANGLKIVDEGYDKTPRIE-KDAFWDDTWVPLYSEASAH 351
QY 347 E 347
DB 352 E 352

Search completed: July 10, 2003, 12:27:10
Job time : 14.0704 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 7.19924 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MTITTSQVAVARLGSFRL.....QLLPSSVEPPLWAPPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	65.6	585	1 VMD2 HUMAN	O76090 homo sapien
2	619.5	26.4	405	1 YS63 CAEEL	O09379 caenorhabdi
3	577	24.6	632	1 Y022 CAEEL	O45435 caenorhabdi
4	573.5	24.4	413	1 Y040 CAEEL	O45435 caenorhabdi
5	571	24.3	499	1 YHDI CAEEL	O23369 caenorhabdi
6	536	22.8	400	1 YV6L CAEEL	O19978 caenorhabdi
7	529	22.5	450	1 YQ83 CAEEL	O17528 caenorhabdi
8	528	22.5	523	1 YQ84 CAEEL	O17529 caenorhabdi
9	512	21.8	459	1 YK78 CAEEL	P34319 caenorhabdi
10	500.5	21.3	420	1 YCBL CAEEL	O94175 caenorhabdi
11	494.5	21.0	456	1 YN4 CAEEL	P34577 caenorhabdi
12	491	20.9	501	1 YSV1 CAEEL	O22566 caenorhabdi
13	450	19.1	513	1 YXAK CAEEL	O21973 caenorhabdi
14	403.5	17.2	884	1 YAVK CAEEL	O17851 caenorhabdi
15	390	16.6	81	1 VMD2 MOUSE	O88670 mus musculu
16	347	14.8	602	1 YH54 CAEEL	O18303 caenorhabdi
17	331	14.1	411	1 YH55 CAEEL	O18304 caenorhabdi
18	285	12.1	434	1 YSWJ CAEEL	O45363 caenorhabdi
19	153	6.5	585	1 ALU5 HUMAN	P39192 homo sapien
20	150	6.4	591	1 ALU8 HUMAN	P39195 homo sapien
21	150	6.4	593	1 ALU7 HUMAN	P39194 homo sapien
22	149	6.3	593	1 ALU6 HUMAN	P39193 homo sapien
23	135	5.7	591	1 ALU1 HUMAN	P39188 homo sapien
24	131	5.6	587	1 ALU2 HUMAN	P39189 homo sapien
25	127.5	5.4	603	1 ALU4 HUMAN	P39191 homo sapien
26	109	4.6	587	1 ALU3 HUMAN	P39190 homo sapien
27	106	4.5	389	1 O85C DROME	O9VHQ6 drosophila
28	104.5	4.4	457	1 SECY CHLTR	P28539 chlamydia t
29	104.5	4.4	613	1 YBID VIBCH	O9KV48 vibrio chol
30	98	4.2	741	1 YBIO ECOLI	P75783 escherichia
31	97	4.1	681	1 CRY1 RATH	O43125 arabidopsis
32	96.5	4.1	3010	1 POLG_HCVUT	Q00269 h genome po
33	95.5	4.1	418	1 YXY1 HUMAN	P49646 homo sapien

RESULT 1

ID	VMD2_HUMAN	STANDARD;	PRT;	585 AA.
AC	O76090; O75904;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bestrophin (Viteliform macular dystrophy protein) (TUI5B).			
GN	VMD2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD.			
RX	MEDLINE=98367043; PubMed=9700209;			
RA	Marquardt A., Stoeck H., Passmore L.A., Kraemer F., Rivera A.,			
RA	Weber B.H.F.;			
RT	"Mutations in a novel gene, VMD2, encoding a protein of unknown			
RT	properties cause juvenile-onset vitelliform macular dystrophy (Best's			
RT	disease).";			
RL	Hum. Mol. Genet. 7:1517-1525(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND B-299.			
RX	MEDLINE=98324772; PubMed=9662395;			
RA	Petrukhin K., Koisti M.J., Bakali B., Li W., Xie G., Marknell T.,			
RA	Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,			
RA	Bergen A.A.B., McGarty-Dugan V., Figueroa D., Austin C.P.,			
RA	Metzker M.L., Caskey C.T., Wadelius C.;			
RT	"Identification of the gene responsible for Best macular dystrophy.";			
RL	Nat. Genet. 19:241-247(1998).			
RN	[3]			
RP	VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.			
RX	MEDLINE=99265978; PubMed=10331951;			
RA	Caldwell G.M., Kakuk L.E., Griesinger I.B., Simpson S.A., Nowak N.J.,			
RA	Small K.W., Maumenee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,			
RA	Ayegari R.;			
RT	"Bestrophin gene mutations in patients with Best vitelliform macular			
RT	dystrophy.";			
RL	Genomics 58:98-101(1999).			
RN	[4]			
RP	VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.			
RX	MEDLINE=99320852; PubMed=10394929;			
RA	Bakali B., Marknell T., Ingvaest S., Koisti M.J., Sandgren O., Li W.,			
RA	Bergen A.A.B., Andreasson S., Rosenberg T., Petrukhin K., Wadelius C.;			
RT	"The mutation spectrum of the bestrophin protein -- functional			
RT	implications";			
RL	Hum. Genet. 104:383-389(1999).			
RN	[5]			
RP	VARIANTS AMD/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.			
RX	MEDLINE=99381534; PubMed=10453731;			
RA	Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,			
RA	Sharma S., Gerrard B., Li W., Metzker M.L., Wadelius C., Caskey C.T.,			
RA	Dean M., Petrukhin K.;			
RT	"Evaluation of the Best disease gene in patients with age-related			
RT	macular degeneration and other maculopathies.";			

Q92777 homo sapien
Q98146 kaposi's sa
P75291 mycoplasma
O82706 salmonella
O82P10 salmonella
Q9PJI1 chlamydia m
P24686 emericella
Q9GZ66 homo sapien
P77172 escherichia
O03368 bos taurus
Q03611 caenorhabdi
P26663 h genome po

ALIGNMENTS

Query Match 65.6%; Score 1541; DB 1; Length 585;
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGERLIFLCYIIRFYRLALTEEQOL 60
 DB 1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGERLIFLCYIIRFYRLALTEEQOL 60

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

QY 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPAEHKQLEKLSLPHNM 180
 DB 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWFVFWFANLGMKAWLGGRIKIRPILLOSLLNEMNTLRTQCGHLAYDWMISIPLVTVQV 240
 DB 181 FWVPWFVFWFANLGMKAWLGGRIKIRPILLOSLLNEMNTLRTQCGHLAYDWMISIPLVTVQV 240

QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290
 DB 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290

RESULT 2
 YS63 CAEEL
 ID YS63 CAEEL STANDARD; PRT; 405 AA.
 AC Q09379;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK675.3 in chromosome II.
 GN ZK675.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sime M.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z46812; CAA86845.1;
 WormPep; ZK675.3; CE01722.
 InterPro; IPR000615; Worm_fam_8.
 Pfam; PF01062; DUF289; 1.
 ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 405 AA; 47835 MW; 0F589D874E9E0E61 CRC64;

Query Match 26.4%; Score 619.5; DB 1; Length 405;
 Best Local Similarity 41.2%; Pred. No. 1.4e-44;
 Matches 121; Conservative 58; Mismatches 102; Indels 13; Gaps 3;

QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGERLIFLCYIIRFYRLALTEEQOL 60
 DB 1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGERLIFLCYIIRFYRLALTEEQOL 60

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
 DB 53 EFKYVLMVDGWTKEIPITFLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

QY 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPAEHKQLEKLSLPHNM 180
 DB 121 GRLLRRLTYRANLGNVILRSYSTAVYKFPSPAQHLVQAGFTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWFVFWFANLGMKAWLGGRIKIRPILLOSLLNEMNTLRTQCGHLAYDWMISIPLVTVQV 240
 DB 181 FWVPWFVFWFANLGMKAWLGGRIKIRPILLOSLLNEMNTLRTQCGHLAYDWMISIPLVTVQV 240

QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290
 DB 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGLKV 290

RESULT 3
 Y022 CAEEL
 ID Y022 CAEEL STANDARD; PRT; 632 AA.
 AC P34672;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK688.2 in chromosome III.
 GN ZK688.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sime M.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L16621; AAA28228.1;
 WormPep; ZK688.2; CE00460.
 InterPro; IPR000615; Worm_fam_8.
 Pfam; PF01062; DUF289; 1.
 ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 73829 MW; 0ABDD1755EF11642 CRC64;

Query Match 24.6%; Score 577; DB 1; Length 632;
 Best Local Similarity 37.9%; Pred. No. 8.3e-41;
 Matches 110; Conservative 65; Mismatches 115; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGERLIFLCYIIRFYRLALTEEQOL 60
 DB 1 MTINYNLAVSTSKWTFLFKLLKWRGSIKAVILAVLVLVGLSVIYRLALNPGQOR 60

QY 61 MPEKLTLYCDSYIOLIPISFVLGFTYVTVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120

```
Db 61 TPERIVQVCDRLSYIPLNFMGFFVTA VVNRWTVLYQIIGFIDNIGLMAAEVVRGRTEQ 120
Qy 121 GRLRLRTRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKQLEKLSLPHNM 180
Db 121 ARMYERNIVRYCELAQVLFRDISTRTRFRFTLDTVVAAGFMWPHKDFDEIQKYKSK 180
Qy 181 FWVPWVFANLSMKAWLGGRIKDPILLQSLLENMNTLRTOCGHLVAYDWISIPLVYTVQV 240
Db 181 YVWPQWAFSLTYEARKKGLIESDYQVQVQDEIKKFRIGLAWICNYDWPVPIPIYPLV 240
Qy 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGLWKV 290
Db 241 CLAVHTYELVCLLARQYVYVSEHADNKTETIDLYFPIMSTLQFIYFGWGMKV 290

RESULT 4
YV4Q_CABEEL STANDARD; PRT; 413 AA.
AC O45435;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Hypothetical protein F32B6.9 in chromosome IV.
GN F32B6.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Baeham V.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z81074; CAB03043.1;
CC WormPep; F32B6.9; CE09864.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
CC SEQUENCE 413 AA; 48965 MW; A66E69A83C78790B CRC64;
CC
CC Query Match 24.4%; Score 573.5; DB 1; Length 413;
CC Best Local Similarity 39.9%; Pred. No. 1e-40;
CC Matches 120; Conservative 53; Mismatches 109; Indels 19; Gaps 5;
Qy 1 MTITYTSQVANARLGSFSLRLLCWRGSIYKLYGELIFLLCYIIRFYRLAL-----T 55
Db 1 MTISYS-----GNVIRILLRWKGSIRWTAKLELLIYLYSVRVFLKGIIDLIDD 52
Qy 56 EQQQL-----MPEKLTLYCDSYIQLIPISFVLGYFTLVVTRWNNQYENLPWDRMLSLVS 111
Db 53 EDDLKMRMPFTFCQCDSYRELPLTLTLLGFFYSNVVNRWQFETLYWPELILSVLC 112
Qy 112 GFVEGKDEQGRLLRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKOL 171
Db 113 TVLHQHDEKSKRRHTIARYLNLALAWRDISSKIRLFPSPVSHSLIESGLLTKKEYQIL 172
Qy 172 EKLSLPH-NMFWVPWVFANLSMKAWLGGRIKDPILLQSLLENMNTLRTOCGHLVAYDWI 230
Db 173 EAMHAENESSRWTTLPHWTQLIMRQVEEHKPTASLNFQVGBELRIFROSLRKLYSYDW 232
Qy 231 SIPLVTVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGLWKV 290
```

```
Db 233 CVPLVTVQVAAALATYSFFFTLGRQLFP-DIETGKELDLVVPVFTVQFLFFVGVFKV 291
Qy 291 G 291
Db 292 G 292

RESULT 5
YHDI_CABEEL STANDARD; PRT; 499 AA.
AC Q23369;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZC518.1 in chromosome IV.
GN ZC518.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Thomas K.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68753; CAA92989.1;
CC WormPep; ZC518.1; CE06601.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
CC SEQUENCE 499 AA; 57229 MW; 4E87725437A5C9DF CRC64;
CC
CC Query Match 24.3%; Score 571; DB 1; Length 499;
CC Best Local Similarity 32.4%; Pred. No. 2e-40;
CC Matches 124; Conservative 70; Mismatches 145; Indels 44; Gaps 6;
Qy 1 MTITYTSQVANARLGSFSLRLLCWRGSIYKLYGELIFLLCYIIRFYRLALTEEQQL 60
Db 1 MTISYTLDSQTNLQSFSLRLLRWGSVMKAVFGQLAVWTVAVFLLISCIYRYMLSPSQD 60
Qy 61 MPEKLTLYCDSYIQL-IPISFVLGYFTLVVTRWNNQYENLPWDRMLSLVSGFVEGKDE 119
Db 61 VPEQLIRYFDNKLDANILPTLLGFFSVFVAVRWGSIILNGICWIDDASLLFATYIRGADE 120
Qy 120 QGRLLRRLIRYANLGNVLILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKQLEKLSLPHN 179
Db 121 ETRVIRRLRVLYLSQALVLRDISMQRVNRKRPDTMTLAASGLMTHHEMDIIDLHIKDPYS 180
Qy 180 MFWVPWVFANLSMKAWLGGRIKDPILLQSLLENMNTLRTOCGHLVAYDWISIPLVYTVQV 239
Db 181 RYMTSIQNSLNLVYECQKKGVDSYLLNKNIVDEIGKFRHGLASLLKYLDWVPVLYPVQV 240
Qy 240 VTVAVYSFELTCLVGRQFL-----NPAKAYPGHELDLVVPVFTFLQFFYVGLWKV----- 290
Db 241 IFLAVRIYFMICLIGRQFIVTGPNS-----GIDLWLPITTMVQFLVYMGWMAEALL 294
Qy 291 -----GLSRALLGWRHGGQGHGQQLLETRMQCQERKVSREBS 328
Db 295 NPLGDEDDDLCEYNIIDKNLITGLSIVDTMWKHDDTGY---SMVEEHM-----AKTPAQKDD 348
Qy 329 QAWRRTPIPATREAEAGESLEP 351
```

```

Dt 349 EFWGIDKIAPLYSMESASERSVHP 371
RESULT 6
ID YV6L CAEEL STANDARD; PRT; 400 AA.
AC Q19578;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein F32G8.4 in chromosome V.
GN F32G8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z73422; CAA97765.1; -
WormPep; B0564.3; CE05177.
InterPro; IPR000615; Worm_fam_8.
PFam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B862E84 CRC64;
-----
Query Match 22.8%; Score 536; DB 1; Length 400;
Best Local Similarity 37.1%; Pred. No. 1.4e-37;
Matches 108; Conservative 58; Mismatches 113; Indels 12; Gaps 4;
QY 1 MTITTSQVANARLGSFRLLCWGRSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTISYDEE-----FSLMLRWGRSINKAVLKOLIGFYIAYIVLAFQWYLLDEKGE 52
QY 61 MPEKLTLYCDSYIQLIPISFVLGFFVTVLVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
DB 53 YFTGWIMWCEIGAQYIPLSFLGFFVSLIVARWQFNCISHPDKMMINVSACLFGNNM 112
QY 121 GRLLRRTLIYRANLGNVILRSVSTAVYKRFPSAQHLVQAGFMTPAHKQLEKLSLPHN 180
DB 113 --VVRQTARSSLSQAIAWSGSVTKLRPPTERHNVASKLWTEEDYDLYNMTDAPHGK 170
QY 181 FWVPWVWFANLGMKAWLGGRIKDPILLQSLNEMNTLTQCGHLYAYDWISIPYTVQV 240
DB 171 WFIPLMIVNLTKKQKQKQ--IDSIQMDMLKQVYSYRDGFAMFLFYDWMIKPLVYTVQV 229
QY 241 TVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQPPFYVGVGLKVG 291
DB 230 AIATYGVFFICLIGRQPKLDQDSME-KEITILFFPIFTFQMLFYLGWLKVG 279
-----
Query Match 22.5%; Score 529; DB 1; Length 450;
Best Local Similarity 37.3%; Pred. No. 6e-37;
Matches 110; Conservative 53; Mismatches 126; Indels 6; Gaps 2;
QY 1 MTITTSQVANARLGSFRLLCWGRSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINYHKEIMTSHPTFFLLPKWKGSIKAVVMTIIFLICYGIIISVIYKTAGMESSQR 60
QY 61 MPEKLTLYCDSYIQLIPISFVLGFFVTVLVTRWNNQYENLPWDRMLSLVSGFVGKDEQ 120
DB 61 TFEISLVRYFDKLSYIPLFLEFVLGFFVTVVNRWTKLYQYTGIFDNLVGLMNCYIRGATEK 120
QY 121 GRLLRRTLIYRANLGNVILRSVSTAVYKRFPSAQHLVQAGFMTPAHKQLEKLSLPHN- 179
DB 121 ARIYRNIMRYCYELVQLVFRDMSMTRRRFPFMTVVAAGFNKHELELYNSYDTKYS 180
QY 180 -----MFVPWVWFANLGMKAWLGGRIKDPILLQSLNEMNTLTQCGHLYAYDWISIPV 235
DB 181 KLGTKYWIPANWALCWTYKARKDGYIESDYFKAQMBGEIRTWRTNIEWVCNVDVPLPLM 240
QY 236 YQVTVVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQPPFYVGVGLKVG 290
DB 241 YPOLVCLAVNLFLVLSIIARQ-LVIEKHQWVDEVDVYFVFMVTFLOPIFYNGWLKV 294
-----
RESULT 8
YOE4 CAEEL
ID YOE4 CAEEL STANDARD; PRT; 523 AA.
AC Q17529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein B0564.4 in chromosome IV.
GN B0564.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U28412; AAL56625.1; -
 DR WormPep; T19C3.1; CE30182.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 501 AA; 57930 MW; 1C56C4C70707FC62 CRC64;

Query Match 20.9%; Score 491; DB 1; Length 501;
 Best Local Similarity 36.0%; Pred. No. 1e-33;
 Matches 105; Conservative 65; Mismatches 118; Indels 4; Gaps 3;

QY 1 MTITTSQVANARLGSFRLLCWGSYKLYGFLIFLLCYIIRFYR--LALTEEQ 60
 DB 23 MTVSQLDVSSGNPLFLRLGRGSIWKSIVGVGLFVWLLFYAIYFYAFYAFSKQLQT 82
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFYVTLVTRWNOYENLPWPDRLMSLVSGFVEGKDEQ 120
 DB 83 VFEISIHDDRMKYLPLFTMLGFFVTIVFERWSALNVNPFIESVALSVALLPKGRE 142
 QY 121 GRLLRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
 DB 143 DELTRAIRYVVLHQLVFRDISMRVRRFPLTKYVDAGFMREQEQLDVLSSVQESSQ 202
 QY 181 -FWPVPWFWFANLSMKAWLGGRIRDPILLOSLLNEMNTLTQCGHLYAYDWISIPLYVTQV 239
 DB 203 TYWVPINWANSALVAHQOKLDQPTAFNNVIFAIKEPRVAMETLIKFDIPIAYPOV 262
 QY 240 VTVAVYSFELTCLVGRQEL-NPAKAYPGHELDLVVPFTFLOFFVVGWLKV 290
 DB 263 VFLAVRVYFAICLVSRQFLISDMKST--QMDWVPVIMTVLEFIFVIGWMKV 312

RESULT 13
 YXAK CAEEL STANDARD; PRT; 513 AA.
 AC Q21973;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein R13.3 in chromosome IV.
 GN R13.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Matthews P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z73105; CAA37442.1; -
 DR WormPep; R13.3; CE06320.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.

KW Hypothetical protein.
 SQ SEQUENCE 513 AA; 58919 MW; 666CB054AD179489 CRC64;

Query Match 19.1%; Score 450; DB 1; Length 513;
 Best Local Similarity 32.9%; Pred. No. 2.9e-30;
 Matches 100; Conservative 57; Mismatches 119; Indels 28; Gaps 5;

QY 1 MTITTSQVANARLGSFRLLCWGSYKLYGFLIFLLCYIIRFYR--LALTEEQ 58
 DB 1 MTVYNLDVSSASIFSLRLQLRWKGSWKYLLKLFMFIIAFITVSSVYRNLITGEKT 60
 QY 59 QLMFEKLTLYCDSYIQLIPISFVLGFYVTLVTRWNOYENLPWPDRLMSLVSGFVEGKD 118
 DB 61 RKINDNFAALPDQNDMDFPLTFMLGFFVTIIVRRNDIPANLGWENTAITVANYIRGTD 120
 QY 119 EGRLLRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPH 178
 DB 121 DRTIRIRNRVIRYVLAQVLVPRDCSIQVRKFPMTMESIVSAGSFQCLGSSATE----- 175
 QY 179 NMFVTPWFWFANLSMKAWLGGRIRDPILLOSLLNEMNTLTQCGHLYAYDWISIPLYVTQ 238
 DB 176 -----YTWSTGLLVDAEAGKIAADLLMNEIGKHIIIEFRKMLALLSNYDWDVPIPLAYPO 259
 QY 239 VTVAVYSFELTCLVGRQEL-NPAKAYPGHELDLVVP-VFTFLOFFVVGWLKV 289
 DB 230 VFLAVRVYFAICLVSRQFLISDMKST-----TVFVMSILOFIFVVGWMK 283
 QY 290 VGLS 293
 DB 284 VAES 287

RESULT 14
 YAVK CAEEL STANDARD; PRT; 884 AA.
 ID YAVK CAEEL
 AC Q17851;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C09B9.3 in chromosome IV.
 GN C09B9.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Bradshaw H.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U50069; AAB37559.2; -
 DR WormPep; C09B9.3; CE29571.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 2.
 DR ProDom; PD002802; Worm_fam_8; 3.
 KW Hypothetical protein; Transmembrane; Repeat.
 FT TRANSMEM 265 285 POTENTIAL.
 FT TRANSMEM 396 418 POTENTIAL.


```

SQ SEQUENCE      884 AA;   102851 MW;   BDB83F9699B8FAID CRC64;

Query Match          17.2%; Score 403.5; DB 1; Length 884;
Best Local Similarity 30.8%; Pred. No. 4.2e-26;
Matches           92; Conservative 63; Mismatches 135; Indels    9; Gaps     5

Qy      1 MTITYTQVANARIGSGSRLLLCWRGSIYKLYGBFLIFLICYIIRFYRLALTEEQOL 60
Db      1 MTISYTYDVATESYFGFPKVLFRWKGSGVKLIHRELFMWLVLYYTTLAIYR-TLDEBRKK 59

Qy      61 MFEXKLTYCDSYIQLIP--ISFVLGGFYTVLVTRRWNYENLPDPDRMLMSLVSGVEGKD 118
Db      60 IFRS---NIEHFNFEPESILTFMLSFFVTITIVQRWNVFTNMGTENAAYASVSFMK-NG 115

Qy      119 EQGLLRRTLTIRYANGLNVILRSVSTAVYKRFPFSACHLVQAQGFWTPAEHRQLEKLSIPH 178
Db      116 EDVERAQTIVRYLVAQILVWRISIKALRRFPFNYESIVTAGFLTKEESTIIIONTDLSY 175

Qy      179 NMFVWPVWVFANLSMKAWLGGRIRDPILOSLINEMNTLTRQCCHLYAYDWISIPLVYTO 238
Db      176 DSSCVPIRAIQVLRHQVRSGNPFSSHVSYYRATWKEVSDFEHLNRVKNVDWVPPLAYPQ 235

Qy      239 VVTVVAYSFFELTCLVGRQFLNPAKAYPGCHELDLVVVPVFTFLOPPFYCGWLKVGSRAL 297
Db      236 VIIFAVELYFVICAFKOYPDLDADDARYVIHYFFIVTFVQFCTCLMGWLKV-AEALL 292

```

RESULT 15	VMD2_MOUSE	STANDARD;	PRT;	81 AA.
AC	C08870;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Bestrophin	(Vitelliform macular dystrophy protein homolog) (Fragment).		
GN	VMD2 OR BMD1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI_TaxID=10090;				
[1]				
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=98324772; Pubmed=9662395;			
RA	Petrukhin K., Koisti M.J., Bakall B., Li W., Xie G., Marknell T.,			
RA	Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,			
RA	Bergen A.A., McGarty-Dugan V., Figueroa D., Austin C.P., Metzker M.L.,			
RA	Caskey C.T., Wadelius C.;			
RT	Identification of the gene responsible for Best macular dystrophy.;			
RT	Nat. Genet. 19:241-247(1998).			
CC	-1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).

```

CC      -----
DR      EMBL; AF057171; AAC64345.1;  -
DR      MGD; MGI:1346332; Vmd2
DR      InterPro; IPR000615; Worm_fam_8.
DR      Pfam; PF01662; DUF289; 1.
DR      ProDom; PD002802; Worm_fam_8; 1.
DR      NON_TER      1
DR      FT      1
DR      NON_TER      81
DR      SEQUENCE      81 AA; 9413 MW; 19FD94AB3D606178 CRC64;
SQ

```

Query Match 16.6%; Score 390; DB 1; Length 81;
Best Local Similarity 86.4%; Pred. No. 3.9e-26;
Matches 70; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Qy 178 HNFVTVVWFANLSKMWLGGRIIRDPIQLQSLLENMTLRTCCGHLVAYDWTISIPLVYT 237

```

Db      1 HNTFWPWVFWANLKMAYLGGRIERTVLLQSLMNEVCTLRTCCQLYADWISIPLVVT 60
Qy      238 QVTVVAVYSFFLTCLVGRQFL 258
Db      61 QVTVVAVYSFFLACLIGKQFL 81

Search completed: July 10, 2003, 12:24:45
Job time : 9.19924 secs

```

Search completed: July 10, 2003, 12:24:45
Job time : 9.19924 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	843.5	35.9	465	11	Q8VCN0	Q8VCN0 mus musculus
2	836	35.6	466	4	Q9NXP0	Q9NXP0 homo sapien
3	763.5	32.5	721	5	Q9V3J6	Q9V3J6 drosophila
4	692	29.4	428	6	Q8WM7	Q8WM7 sus scrofa
5	685	29.1	535	5	Q9VUM7	Q9VUM7 drosophila
6	610.5	26.0	612	5	O17206	O17206 caenorhabdi
7	590	25.1	557	5	O17205	O17205 caenorhabdi
8	559	23.8	584	5	Q18303	Q18303 caenorhabdi
9	543.5	22.1	327	5	Q9VUM6	Q9VUM6 drosophila
10	535	22.8	1447	5	O16779	O16779 caenorhabdi
11	493	21.0	731	5	Q9VRW4	Q9VRW4 drosophila
12	476.5	20.3	512	5	O62095	O62095 caenorhabdi
13	465	19.8	551	5	Q9NA59	Q9NA59 caenorhabdi
14	391	16.6	551	5	Q965X4	Q965X4 caenorhabdi
15	382	16.3	400	5	O17674	O17674 caenorhabdi
16	351.5	15.0	444	5	O17671	O17671 caenorhabdi

```
Qy 230 ISIPLVYTVVTVVAVYSPFLTCLVGRQFLNPAKAYEGHDLVVPVFTLQFFVYVGMWK 289
Db 187 ISIPLVYTVVTVVAVYSPFLTCLVGRQFLNPAKAYEGHDLVVPVFTLQFFVYVGMWK 246
Qy 290 VGLSPALLGWRHGRQHGGQ---LLETRMQCQKVRVSSQA-----WW-----RTP 335
Db 247 V--ASQLN-PEGEDDDDFETNFDNRNFQVSMVLAVDEMYDYLAMLEKDLVYDAEAP 303
Qy 336 VIPATR-----EAEAGESLEPGR-----RRLWMOSSST 364
Db 304 YTAATAPLLOQPSFOGSTFDIALAKEDMQFLRDGVDGPGLVGHGDFLQRLLPAGAGSVG 363
Qy 365 PLEMMMLIR-----PTGLSTGICRCPCWLMWRCRTCLGMSRTC-----TGISPSH 411
Db 364 PLGRRLSLRLKRNKSCVSEASTAASCG-----CAGAAGGGVGECCGDLPLDPSL 412
Qy 412 SPPTQLLPSSVEPPLWAPPST 433
Db 413 REP-ELEPPACPEPPAPIPGPT 433

RESULT 2
Q9NXP0 PRELIMINARY; PRT; 466 AA.
AC Q9NXP0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CDNA FJ20132 fis, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Oiyashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000139; BAA0970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

Query Match 35.6%; Score 836; DB 4; Length 466;
Best Local Similarity 63.5%; Pred. No. 8.6e-65;
Matches 153; Conservative 35; Mismatches 53; Indels 0; Gaps 0;

Qy 50 YRLALTEQQLMFEKLTLYCDSYIOLPISFVLGFTVTVTRWNNQYENLPKPOLMSL 109
Db 7 YRFVTEGQKRYFEKLVICDQYASLIPVSFVLGFTVTVTRWNNQYENLPKPOLMSL 66
Qy 110 VSGFVEGDEQGLRLRLTIRYANLGNVLILRSVAVYKRPFSAGHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDGRGLYRRLTMRVYAGLSAVILRSVAVYKRPFTIDHVGAGFMTPEERK 126
Qy 170 QLEKLSLPHNFWPVPWVPANLSMKAWLGGRTRDPILLOSILNEMNTLTQCGHLYAYDW 229
Db 127 KPENLNSYNKYVPCVWFPSNLAAQARREGIRDNLSALKLLLELNLFVFGKCMFLFYDW 186
Qy 230 ISIPLVYTVVTVVAVYSPFLTCLVGRQFLNPAKAYEGHDLVVPVFTLQFFVYVGMWK 289
Db 187 ISIPLVYTVVTVVAVYSPFLTCLVGRQFLNPAKAYEGHDLVVPVFTLQFFVYVGMWK 246
Qy 290 V 290
Db 247 V 247
```

```
RESULT 3
Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-WAY-2000 (TReMBLrel. 13, Created)
DT 01-WAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6264 protein (LP05915P).
GN BEST OR CG6264.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA April J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Taveanli B.C., Pappu K., Mardon G.;
RT "dbest, a Drosophila Homolog of Human Bestrophin."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003686; AAF54503.1; -.
DR EMBL; AF218817; AAF32327.1; -.
DR EMBL; AF061546; AAL29094.1; -.
DR DR
```



```

Db 1 MTVSATVATCSHFQCFWKLMEWRAIYKIIVDLALFLSCFYEMAVIYRYALRDVDK 60
Qy 60 LMFEKLTLYCDSYIOLIPISFVLGFTLVTRWNNQYENLPWDRMLSLVSGFVEGKDE 119
Db 61 PVFEDIVMCHSYNSNLIPLSFVLGFTVGIIRWNNQYITVPDPPLAVYVSALVRGQDE 120
Qy 120 QGRLLRRLTIYANIGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAHKKOLEKLSL--- 176
Db 121 HGRLLRRLTIYANIGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAHKKOLEKLSL--- 180
Qy 177 PHNMFVFWPFWANLKMALGGRIIDPILQSLNENMTLRTQCGHLYAYVDWISIPLYVTV 236
Db 181 KHPYVWPIVAAISIVTRAKGRIWDPSLKSMDLKNFRACGNMLIHVDITISVPLVY 240
Qy 237 TQVTVVAYSPFLICLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFFVVGWLVK 290
Db 241 TQVTVLAVYSYFVASIFGHOWIDRIDIKHNNIVSYYPFLFSTLEPFFFMGLKV 294

RESULT 6
Ol7206 PRELIMINARY; PRT; 612 AA.
AC Ol7206;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE COIB12.3 protein.
GN COIB12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025458; AAB70976.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 612 AA; 71031 MW; DFBB43916541DD44 CRC64;

Query Match 26.0%; Score 610.5; DB 5; Length 612;
Best Local Similarity 32.7%; Pred. No. 5.7e-45;
Matches 130; Conservative Indels 47; Gaps 6;

Qy 1 MTITYSQVANARLGSFSLLCWRGSIYKLLYCEFLIFLLCYIIRFYRLALTEEQQL 60

```

```

Db 1 MTITYSLDVASSFFCLYKLLFRWKGSIKSVWAEVLVWLCLYAVLSVYVYRCLLTMKORA 60
Qy 61 MFEKLTLYCDSYIOLIPISFVLGFTLVTRWNNQYENLPWDRMLSLVSGFVEGKDEQ 120
Db 61 TFDLCLFFDTYSNFIDITFALGFYSAVTRWQIIFDNIGMIDTPCLMIQYIKGETER 120
Qy 121 GRLRLRLLIYANIGNVLIILRSVSTAVYKRPSPSAHLVQAGFMTPAHKKOLEKLSLPHNM 180
Db 121 AKCVRNRCIRYSILTOAMVYRDVAASVRKRPFTFNHLVTLAGLMTKEKNAEFESIPSPHAK 180
Qy 181 FWPVWFWANLKMALGGRIIDPILQSLNENMTLRTQCGHLYAYVDWISIPLYVTVV 240
Db 181 YQPMHMLFSLMITARDEGMISDIIYVDLMKMRQFRVNLSTLFDWVPVPLVTVV 240
Qy 241 TVAVYSFPLICLVGRQFLNPAKAYPGHLDLVVPVFTFLOPFFVVGWLVKGLSRALL 297
Db 241 HLAVRSYFLIALFGQYLHPESNLNDPKQITIDLYVPIMSLQFIFFIGWKKV--AEVLL 298
Qy 298 -----GWRHGQRGHGQQLLETRMQCQERKVSRESSQAWWTFVPIPATREAEA 345
Db 299 NPLGHEDDDDFEENW-----ILDRNLQ-----VGLMVVDYAYNRYPTL----- 335

RESULT 7
Ol7205 PRELIMINARY; PRT; 557 AA.
AC Ol7205;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE COIB12.5 protein.
GN COIB12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Scheet P., Maggi L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025458; AAB70976.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 557 AA; 64076 MW; 8A68BBB1436D0331 CRC64;

```

```

Query Match      25.1%; Score 590; DB 5; Length 557;
Best Local Similarity 33.8%; Pred. No. 3.2e-43;
Matches 125; Conservative 79; Mismatches 142; Indels 24; Gaps 5;

QY 1 MTITTSQVANARLGSFLLLCWRGSIYKLLYGEFLPLCYIIRFYRLALTEEQQL 60
DB 1 MTAVSLDVATSGLFTQIKVLLRWKGSVMKSIWSELLIWLCCSILSVIRILLNKAQRE 60
QY 61 MFEKLTLYCDSYQILPISFVLGFTVTVVTRWNOYENLPWDRMLSLVSGFVEKDEQ 120
DB 61 VPEQLCTPDTFSPFVTFMGLGFTVIVNWKVFDNVGMDTSALTIAQYIRTSK 120
QY 121 GLLRLTIRYANLGNVLLRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKOLEKLSLPHNM 180
DB 121 ARLIRNCRVYRIVAVQWVFRDVSFAIRRRFFTKIHLIGAGLITDELTEFADITSPQSK 180
QY 181 FWVPWFWFANLSMKAWLGRIRDPILLQSLNEMNTLTCQGHLYAYDWISIPLVYTVQV 240
DB 181 YWQPIQWLFSLVTVADEGLIADSVLYDLIDKRRFRKIINLVIFDMVPIPLVYTVQV 240
QY 241 TVAVYSFELTCLVGRQPLNPAKAYPG--HELDLVVPVFTFLOFFFYVGLKVLGRALIG 298
DB 241 NLAVRTYFVALFGRLNNNNIFGAKWKIDYFPINTSLQIVFVGLKV--SEWML- 297
QY 299 WRHGORGHGOQLLETRMOCQER-KVSRVSSQAMWRTVPVPATREAEAGESLEPGRRRLW 357
DB 298 ---NPLGDEDEDFETWIIERNLQVGVAVDQVGYPTL-----KRDFF 339
QY 358 WQSSSSTPLE 367
DB 340 WEDETPQTM 349

RESULT 8
Q18303 Q18303 PRELIMINARY; PRT; 584 AA.
AC Q18303
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DE C29F4.2 protein.
GN C29F4.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP Kershaw J.K.;
RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RL SEQUENCE FROM N.A.
RN [2]
RP Kershaw J.K.;
RA none;
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z68335; CAA92730.1; -;
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 584 AA; 68252 MW; EGAB86A4EEB21BCC CRC64;

Query Match      23.8%; Score 559; DB 5; Length 584;
Best Local Similarity 36.9%; Pred. No. 1.7e-40;
Matches 101; Conservative 64; Mismatches 109; Indels 0; Gaps 0;

QY 19 RLLLCWRGSIYKLLYGEFLPLCYIIRFYRLALTEEQQLMFEKLTLYCDSYQIPI 78
DB 63 KLIFKWKGLSKWAIYLDLIVWCFCYATSVIYRVALDRSQDTERFMQCNRLDIPI 122
QY 79 SFVLGFTVTVVTRWNOYENLPWDRMLSLVSGFVEKDEQGRLLRRLTIRYANLGNV 138

123 NFMGLGFFVTTVINRWMTQPANLGMIDNIALFTSMYLSGDERGRILLRSIVRCMWSQTM 182
139 ILRSVSTAVYKRPSPSAQHLVOAGFWTPAEHKOLEKLSLPHNMFWVPWVPMANLSMKAWLG 198
183 VFRDIHIGVRKFPPTLETVAAGIMTSSBLKTKYNEVSRYAKTWLGFNMTFLLNREARRE 242
199 GRIRDPIQLQSLNEMNTLTCQGHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQPL 258
243 GRIESAYTQNAEERTFRSGLSLIWTYDWPILMYPOLVFWMAHCYYLVLCVLSQPV 302
259 NPAKAYPGHELDLVVPVFTFLOFFFYVGLKVL 292
303 INSDAVNTTEIDLGVPFMTIIFIFVGLKVKAM 336

RESULT 9
Q9VUM6 Q9VUM6 PRELIMINARY; PRT; 327 AA.
AC Q9VUM6
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DE 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE CG7259 protein.
GN CG7259
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
EMBL; AE003531; AAF49649.1; -;

```

```

DR FlyBase; FBgn0036491; CG7259.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 327 AA; 38594 MW; BE7CE3BC05370F46 CRC64;

Query Match      23.1%; Score 543.5; DB 5; Length 327;
Best Local Similarity 44.3%; Pred. No. 28-39;
Matches 101; Conservative 48; Mismatches 72; Indels 7; Gaps 2;

QY 68 YCDSYLIQIPISFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKDSQGLLRRT 127
DB 3 YCKKNSLIPISFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKDSQGLLRRT 62

QY 128 LIRYANLGNVILRSVSTAVYKRPSPSAOHLVQAGFMTPAEHKQLEK-----SLPHNMF 182
DB 63 ILRYVCLQVIVFTWISPRKRRFPYTIQIEAGFLENEKKIEAMDQAPSPYKHH--W 120

QY 183 VPWVFWFANLSMKAWLGGRIKRDPIILQSLNEMNTLTQCGHLYAYDWISIPLYTVQVTV 242
DB 121 MPVWAAISVMPARRENKIRDDYAVKTIIDELNLRGNCGLFYDYDWISVPLVYTVQVTV 180

QY 243 AVYSFELCLVGRQFLNPAKAYPGHELDLVVPVTFLOPFYVGMVKV 290
DB 181 ATYSFELSVLQQQNESHSDGTRIRRWFPILTVLQFFTFMGMLKV 228

RESULT 10
O16779 PRELIMINARY; PRT; 1447 AA.
AC O16779; O16781;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical.164.2 kDa protein.
GN T21d12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Woessner J.;
RT "The sequence of C. elegans cosmid T21d12."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016687; AAK72064.1; -.
DR HSSP; P56276; 1TLK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR000815; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.

DR Pfam; PF00047; ig; 3.
DR Pfam; PF00560; LRR; 12.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICRPT.
DR ProDom; PD002802; Worm_fam_8; 1.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 2.
DR SMART; SM00410; IG like; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TIP; 13.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 1447 AA; 164168 MW; 4BC1FE38E9E85B5 CRC64;

Query Match      22.8%; Score 535; DB 5; Length 1447;
Best Local Similarity 37.7%; Pred. No. 68-38;
Matches 115; Conservative 61; Mismatches 111; Indels 18; Gaps 5;

QY 1 MTITYTSQVANARLGSFRLCLCMRGSYKLYGFLFLCLCYIIRFYR--LALTEEQ 58
DB 879 MTVSYNSDVSTSLTAFRLQLRWGWSVWRELFIFFCFALITSYRTNYFLSDEQ 938

QY 59 QLMPEKFLYCDYSYLIQIPISFVLGVYTLVTRWNOYENLPWDRMLSVSGFVEKGD 118
DB 939 RVFMDNFAALFDQTLNLYIPTLTMGLGFFVTIIVGRWNDIFLNIGWVDNTALLIATYIRGG 998

QY 119 EQRLLRRTLRVYANLGNVILRSVSTAVYKRPSPSAOHLVQAGFMTPAEHKQLEKSL-P 177
DB 999 EKSRIWRTALRYLVLTQVLPFRDISLVVKGFFPILTVVAGSFMLESEKRYAALSGLK 1058

QY 178 HNPMPVWVFWFANLSMKAWLGGRIKRDPIILQSLNEMNTLTQCGHLYAYDWISIPLYVT 237
DB 1059 YPKYELPIQWCFSLYDARAQCKIGADYVNLKSVGDFRKGGLQLNDFWVPIPLVYP 1118

QY 238 QVTVVAVVSFFLTVGRQFL-----NPAKAYPGHELDLVVP-VFTFLOPFYVGMVL 288
DB 1119 QVVFVAVRIYFPLCLIRQSVLIDGKPKDNPVYP-----FVPFLMTSLOPFVYVGM 1172

QY 289 KVGLS 293
DB 1173 KVABS 1177

RESULT 11
O9VRW4 PRELIMINARY; PRT; 731 AA.
AC O9VRW4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG10173 protein.
GN CG10173.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,
RA George R.A., Lewis S.E., Yandell M.D., Zhang Q., Chen L.X.,
RA Sutton G.G., Wortman J.R., Blazew R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```


RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Stiden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	132	133	61	191	121	251	181
QY	78	133	61	191	121	251	181	
Db	1							
QY	133							
Db	61							
QY	191							
Db	121							
QY	251							
Db	181							
RESULT 12								
O62095								
ID	O62095	PRELIMINARY;	PRT;	512 AA.				
AC	O62095;							
DT	01-AUG-1998 (TrEMBLrel. 07, Created)							
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)							
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)							
DE	C37A5.1 protein.							
GN	C37A5.1.							
OS	Caenorhabditis elegans.							
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;							
OC	Rhabditidae; Pelodierinae; Caenorhabditis.							
OX	NCBI TaxID=6239;							
XP	[1]							
RRP	SEQUENCE FROM N.A.							
RRP	White S.;							
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.							

DR EMBL; Z82095; CAB05710.1; JOINED.
DR EMBL; Z82095; CAB05029.1; --
DR EMBL; Z83221; CAB05029.1; JOINED.
DR InterPro; IPR000615; Worm_fam 8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam 8; 1.
SQ SEQUENCE 400 AA; 46827 MW; D1CDF601E25E2D01 CRC64;

Query Match 16.3%; Score 382; DB 5; Length 400;
Best Local Similarity 31.2%; Pred. No. 3.1e-25;
Matches 100; Conservative 53; Mismatches 113; Indels 54; Gaps 8;
QY 1 MTITYTSQVANARLGSFRLLCWRGSIYKLYGEFLPLCYIIRPIYRLAL-----T 55
Db ||||| :
1 MTITYYRAVSTESISNPLKIIIRHWHGSLAKSVQEEFLWLPFYVVCAYRCVLP 60
QY 56 BEQQLMFEKLTLYCDSYIQLIPISFVLGXYTLVTRWNNQYENLPWPORLMSLVSGFVE 115
Db :
61 DQFKKFVEILSTHQDNY---IPLEFVLGPFVTVIDRWKSFQNIPIESCAFAVSAALP 117
QY 116 GKDEQGRU-----LRRTLIYANLGNVLILRSVSTAVYKRPSPQAHLVOAGFM 163
Db ||| :
118 G---HGALEVSAGMDTKLTARTTIIRYLVLSQLLFREFSVKVKCRFVDRSLVDSKFL 174
QY 164 TPAEHKQL--EKLSLPHNMFWPWWWFANL-----SMKAWLGGRIRDPILLQSL 210
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
175 TENELTELSEELKTKRYSIILPINWAFSILRTEKLNPNQPMNAN----- 220
QY 211 LNMENLRTQCGHLYADWISIPLVYTVQVTVAVYSPFLTCLVGRQFLNPAKAYPCHELD 270
Db :
221 -NVINDWQVKLTLLRNGDFIPIPIAYPOAVFLAVRFYFLVCLFTRQHLDS---DHHAID 276
QY 271 LVVPVFTFLQFPFYGVNLKV 290
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Search completed: July 10, 2003, 12:26:19
Job time : 25.8129 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 34.021 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTKVANARLGSFSSL.....EHAESYPRDEAGTKPVLYE 551

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2944	100.0	551	20 AAY29955	Mouse CGICE protei
2	1899	64.5	585	20 AAY29953	Human CGICE long f
3	1297	44.1	435	20 AAY29954	Human CGICE short
4	961	32.6	721	22 ABB62213	Drosophila melanog
5	858.5	29.2	535	22 ABB63121	Drosophila melanog
6	752	25.5	327	22 ABB63135	Drosophila melanog
7	728	24.7	368	22 ABB60609	Novel human diagno
8	714.5	24.3	731	22 ABB64914	Drosophila melanog
9	641	21.8	314	22 AAB70085	Human secreted pro
10	534	18.1	251	22 AAB70067	Human secreted pro

11	534	18.1	251	23 ABB65506	Human albumin fusi
12	444	15.1	261	20 AAW95345	Human adult retina
13	374	12.7	101	22 ABB14768	Human nervous syst
14	313	10.6	701	22 ABB08608	Novel human diagno
15	232	7.9	99	22 AAU22652	Novel human colon
16	232	7.9	99	22 AAW92658	Human digestive sy
17	207.5	7.0	377	22 ABB08411	Novel human diagno
18	165	5.6	30	20 AAY29956	Human CGICE partia
19	122.5	4.2	188	23 ABB92888	Human polypeptide
20	122	4.1	47	22 ABB08607	Novel human diagno
21	119	4.0	30	20 AAY29968	C. elegans CGICE r
22	119	4.0	30	20 AAY29971	C. elegans CGICE r
23	115	3.9	30	20 AAY29969	C. elegans CGICE r
24	110	3.7	384	21 AAB26452	Drosophila melanog
25	110	3.7	384	22 ABB66539	Drosophila melanog
26	109.5	3.7	1816	21 AAY95440	Caenorhabditis ele
27	107.5	3.7	470	23 AAO14194	Human transporter
28	104.5	3.5	389	21 AAB20926	Drosophila odorant
29	102.5	3.5	1522	20 AAW99302	Human BA13 protein
30	102.5	3.5	1522	21 AAB23601	Human secreted pro
31	102	3.5	30	20 AAY29970	C. elegans CGICE r
32	101.5	3.4	701	22 ABB25778	Novel human diagno
33	99.5	3.4	361	23 ABB26907	Streptococcus poly
34	98.5	3.4	361	23 ABB23762	Streptococcus poly
35	99.5	3.4	855	13 AAR29815	S receptor kinase
36	99	3.4	477	21 AAG27305	Arabidopsis thalia
37	99	3.4	477	21 AAG27304	Arabidopsis thalia
38	99	3.4	533	21 AAG27303	Arabidopsis thalia
39	97	3.3	642	22 ABB59242	Drosophila melanog
40	96.5	3.3	567	22 AAB94166	Human protein sequ
41	96.5	3.3	839	20 AAY30152	A human vanilloid
42	96.5	3.3	839	20 AAY30153	A partial human va
43	96.5	3.3	839	20 AAY30155	A human vanilloid
44	96.5	3.3	839	20 AAY06558	Human capsaicin re
45	96.5	3.3	839	21 AAB332127	Human vanilloid re

ALIGNMENTS

RESULT 1

AA29955

ID AAY29955 standard; Protein; 551 AA.

XX

AC AAY29955;

XX

DT 22-NOV-1999 (first entry)

XX

DE Mouse CGICE protein sequence.

XX

KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

KW BMD; age-related macular dystrophy.

XX

OS Mus sp.

XX

PN WO9943695-A1.

XX

PD 02-SEP-1999.

XX

PF 22-FEB-1999; 99WO-US03790.

XX

PR 25-FEB-1998; 98US-0075941.

XX

PR 18-DEC-1998; 98US-0112926.

XX

PA (MERI) MERCK & CO INC.

XX

PA (UYUP-) UNIV UPPSALA.

XX

PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;

 XX || DR | WPI, 1999-540560/45. | |
DR	N-FSDB; AAZ21229.	
XX		
PT	Human and mouse polynucleotides encoding CGICE polypeptides	

Db 417 RESLLHEGLPKNHKAQKQVRQEDNKAWLKAVDAFKSGPLQYQRPYYAPQTPLSPTP 476
Qy 469 --HPTQAPSS--SDTG-----DGPSTDYQICHMKKKT 499
Db 477 MFFPLEPAPSKLHVSVTGIDTKSLKTSVSGAKSKFELLSSDGMALMEHPVSOVRRKT 536
Qy 500 VEFNL-NTPESPTEHLQRRDDOMSTNIOALMKHAESV---PYDEA 543
Db 537 VEFNLDMPEIPENHLKE-PLQSQSPNTNHTTLKOHMDPYWALENDEA 583

RESULT 3
AAY29954
ID AAY29954 standard; Protein; 435 AA.
XX AC AAY29954;
XX DT 22-NOV-1999 (first entry)
XX DE Human CGICE short form protein sequence.
XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX OS Homo sapiens.
XX PN WO9943695-A1.
XX PD 02-SEP-1999.
XX PF 22-FEB-1999; 99WO-US03790.
XX PR 25-FEB-1998; 98US-0075941.
XX PR 18-DEC-1998; 98US-0112926.
XX PA (MERI) MERCK & CO INC.
PA (UTUP-) UNIV UPPSALA.
XX PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX WPI; 1999-540560/45.
XX DR N-PSDB; AAZ21228.
XX PT Human and mouse polynucleotides encoding CGICE polypeptides -
XX PS Claim 7; Fig 5; 67pp; English.
XX CC The present sequence represents the human CGICE protein. When the CGICE
CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
CC Polynucleotides encoding CGICE are useful for diagnosing whether a
CC patient carries a mutation in the CGICE gene. Normal and mutated
CC CGICE proteins are useful for identifying activators and/or inhibitors
CC of these proteins, in order to treat BMD. The CGICE gene offers a
CC simpler and cheaper method of diagnosing BMD without the need for the
CC presence of the patient. The gene may also be useful to discovering
CC the genetic cause of age-related macular dystrophy.
XX SQ Sequence 435 AA;
Query Match 44.1%; Score 1297; DB 20; Length 435;
Best Local Similarity 82.4%; Pred. No. 1.3e-125;
Matches 239; Conservative 23; Mismatches 28; Indels 0; Gaps 0;
Qy 1 MTITTYNKVANARLGSFSLLCWRGSIYKLYGBFLVFIFLYSTIRGLYRMVLSDDQL 60
Db 1 MTITTSQVANARLGSFSLLCWRGSIYKLYGBFLVFIFLYSTIRGLYRMVLSDDQL 60
Qy 61 LFEKALYCDYSYIQLIPISFVLGFTVTLVSRWMSQYENLPWDRMLMQVSSFVGGKDE 120
Db 61 MPEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLMSLVSGFVGGKDEQ 120
Qy 121 GELLRTLRIRYANLGNVLLRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Db 121 GELLRTLRIRYANLGNVLLRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Qy 181 FWVPWVWFANLSTKAYIGGRIRDTVLQSLANEVCTLTQCGQLYAYDWISIPLYTQVV 240
Db 181 FWVPWVWFANLSTKAYIGGRIRDTVLQSLANEVCTLTQCGHLAYAYDWISIPLYTQVV 240
Qy 241 TVAVTSFELACLIGROFLNPNKDYPGHEMDLVVPVFTILQFLFYNGWLKV 290
Db 241 TVAVTSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFFYVGNLKV 290

RESULT 4
ABB62213
ID ABB62213 standard; Protein; 721 AA.
XX AC ABB62213;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 13431.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL06316.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX PS Disclosure; SEQ ID NO 13431; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 721 AA;
Query Match 32.6%; Score 961; DB 22; Length 721;
Best Local Similarity 50.4%; Pred. No. 2.4e-90;
Matches 191; Conservative 70; Mismatches 108; Indels 10; Gaps 6;
Qy 1 MTITTYNKVANAR-LGSPSSLLLCWRGSIYKLYGFLVFIFLYSTIRGLYRMVLSDDQ 59
Db 1 MTITTYGVAICRGPGCFKLLLRWRSIYKLYWLDLAFLLTYIYAINNVYRFLNPAQK 60
Qy 60 LFEKALYCDYSYIQLIPISFVLGFTVTLVSRWMSQYENLPWDRMLMQVSSFVGGKDE 119
Db 61 ETEFAIVQCDYSYRELILPLSFVLGFTVTRWNNQTSIPWDPDIAVFNSSNVHGQDE 120

cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 327 AA;

Query Match 25.5%; Score 752; DB 22; Length 327;
Best Local Similarity 47.7%; Pred. No. 3.9e-69;
Matches 143; Conservative 54; Mismatches 97; Indels 6; Gaps 3;

QY 68 YDSYITQLIPISVLGYVTLVSVRMSQYENLPWDLMIQVSVFEGKDEGRLLRRT 127
DB 3 YCEKNGSLIPSLSVLGFVFRVNRWMDQYTTIPWPDGIALISTSHGSDDRARVWRT 62

QY 128 LIRYAILGVLLIRLSISTSVYKRPPTLHLVLGAFTHGHEKLOKLGLPHNTF---WVP 184
DB 63 ILRYVCLQVIVFTMTISPRVRRFPPTYQTIIAGFLLENEKKIIEAMDQAFPSYKHWMP 122

QY 185 WYFANLSMKAYLGGRIKRDVTLVLSLMNEVCTLRTOGOLYAYDWISIPLVYQVTVAV 244
DB 123 IVWASIVNRARENKIRDDYAVKTIIDELNQRGCGFLLYDYDWISVPLVYQVTVAT 182

QY 245 YSFFLAQLGROFLNPNKDYPGHEMDLVVPVFTILOFLYMGWLKVAEOLINPFGEDDD 304
DB 183 YSFFLPSVLGQWQNESHSDGTRIRRWFPILTVLQFFYMGWLKVAETLINPFGEDDD 242

QY 305 FETNWIIDRLNQLVSLVSGMHQNLPPMERDMYWNAAAP-QPPYTAASARRSHSPMGST 363
DB 243 FELNWIIDRLNQLVSYCIVDMHQEHPDLVKDQWEEVFPNEIPY--AQPQKQNPAAST 300

RESULT 7
ABG08609
ID ABG08609 standard; Protein; 368 AA.
AC ABG08609;
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #8600.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PP 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
DR WPI; 2001-639362/73.
DR N-PSDB; AAS72796.
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PS Claim 20; SEQ ID No 38968; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 368 AA;

Query Match 24.7%; Score 728; DB 22; Length 368;
Best Local Similarity 60.5%; Pred. No. 1.4e-66;
Matches 135; Conservative 30; Mismatches 54; Indels 4; Gaps 1;

QY 162 FMTGHEHQLQKLGPHNTFWVPWFANLSMKAYLGGRIKRDVTLVLSLMNEVCTLRTOC 221
DB 1 FMTTDERKLFNHLKSPHLKYVVPFWFVFNLAATKARNEGRIKRDVTLVLSLMNEVCTLRTOC 60

QY 222 GOLYAYDWISIPLVYQVTVAVYVFFLAQLGROFLNPNKDYPGHEMDLVVPVFTILOP 281
DB 61 SLLFGYDWGIPLVYQVTVAVYVFFLAQLGROFLNPNKDYPGHEMDLVVPVFTILOP 120

QY 282 LFYMGWLKVAEOLINPFGEDDDPETNWIIDRLNQLVSLVSGMHQNLPPMERDMYWNAA 341
DB 121 PFYAGWLKVAEOLINPFGEDDDPETNWIIDRLNQLVSLVSGMHQNLPPMERDMYWNAA 180

QY 342 APQPYTAASARRSHSPMGSTFNISLKKEDLE----LWSKEE 380
DB 181 AARPEYTLAAADYCIPTSPFLGSTVQMLSGSDPPDEBWLWDYBK 223

RESULT 8
ABBS64914
ID ABBS64914 standard; Protein; 731 AA.
AC ABBS64914;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 21534.
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
OS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
PP 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;

KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vaccine;
 KW antialzheimers; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.

OS Homo sapiens.

PN WO200112776-A2.

XX 22-FEB-2001.

XX 15-AUG-2000; 2000WO-US22350.

XX 16-AUG-1999; 99US-0148759.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;

PI WPI: 2001-244245/25.

DR N-PSDB; AAF76848.

XX Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -

PS Claim 11; Page 362-363; 380pp; English.

XX The present sequence is one of 18 novel human secreted proteins. The
 CC nucleic acids encoding the proteins and the proteins themselves may be
 CC used in the prevention, diagnosis and treatment of diseases including
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
 CC and human immunodeficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration
 CC and/or chemotaxis. The nucleic acid molecules may also be used as DNA
 CC probes in diagnostic assays to detect and quantitate the presence of
 CC similar nucleic acid sequences in samples. The polypeptides may also be
 CC used as antigens in the production of antibodies and in assays to
 CC identify modulators of protein expression and activity.

XX Sequence 251 AA;

Query Match 18.1%; Score 534; DB 22; Length 251;

Best Local Similarity 55.9%; Pred No. 1.2e-46;

Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LYAVDWISIPLVYTVQVTVAVYSPFLACIGRQFLNPNKD-----YPGHE----- 268

Db 2 LFHYDWISIPLVYTVQVTVAVYSPFALSILVGRQFVEPEAGAKPKQLKPGQEPAPALGD 61

QY 269 MDLVVVPVFTILOFLFYMGWLKVAEQLINPFGEDDDDFETNWIIDRNQLVSLSDVMGHON 328

Db 62 PDMVPLTLLQOFFFYAGWLKVAEQIINPFGEDDDDFETNQLIDRNQLVSLSDVMYQON 121

QY 329 LPPMERDMYNEAAPQPPYT--AASARRSRHSGFMSTFNISLKKEDELWSKEEADTKKE 387

Db 122 LPPAEKQYWDQPPQPPYTVATAESLRPFLSGTFLNLRMSDDPEQSLQVFASPGSGRP 181

QY 388 SGYSST--IGCFGLG 400

Db 182 APAQATLLGRFLGV 196

RESULT 11

ABG5506

ID ABG5506 standard; Protein; 251 AA.

XX

AC ABG5506;

XX 27-AUG-2002 (first entry)

XX Human albumin fusion protein #2181.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antifertility; antinflammatory; antiulcer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.

OS Homo sapiens.

OS Synthetic.

XX WO200177137-A1.

XX 18-OCT-2001.

XX 12-APR-2001; 2001WO-US11988.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;

PI WPI: 2002-010886/01.

XX New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -

XX Claim 1; Page 2075; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia) and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG53326-ABG5518 represent albumin
 CC fusion proteins of the invention.

XX Sequence 251 AA;

Query Match 18.1%; Score 534; DB 23; Length 251;

Best Local Similarity 55.9%; Pred No. 1.2e-46;

Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LYAVDWISIPLVYTVQVTVAVYSPFLACIGRQFLNPNKD-----YPGHE----- 268

Db 2 LFHYDWISIPLVYTVQVTVAVYSPFALSILVGRQFVEPEAGAKPKQLKPGQEPAPALGD 61

QY 269 MDLVVVPVFTILOFLFYMGWLKVAEQLINPFGEDDDDFETNWIIDRNQLVSLSDVMGHON 328

Db 62 PDMVPLTLLQOFFFYAGWLKVAEQIINPFGEDDDDFETNQLIDRNQLVSLSDVMYQON 121

QY 329 LPPMERDMYNEAAPQPPYT--AASARRSRHSGFMSTFNISLKKEDELWSKEEADTKKE 387

Db 122 LPPAEKQYWDQPPQPPYTVATAESLRPFLSGTFLNLRMSDDPEQSLQVFASPGSGRP 181

QY 388 SGYSST--IGCFGLG 400

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-541565/60.
N-PSDB; ABA11094.
Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -
Claim 11; SEQ ID NO 3425; 1701pp + Sequence Listing; English.
The invention relates to novel genes (ABA11004-ABA21534) and proteins (AB114678-AB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 101 AA;

Query Match 12.7%; Score 374; DB 22; Length 101;
Best Local Similarity 74.5%; Pred. NO. 1.3e-30;
Matches 70; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 291 AEQLINPEGDDDDFFETNWIIDRLNQLVSLSDVGHQNLPMERDMYNRAAPQPPYTA 350
DB 1 AEQLINPEGDDDDFFETNWIIDRLNQLVSLSDVGHQNLPMERDMYNRAAPQPPYTA 60
QY 351 SARRRRHSMGSTFNISLKEDLELWSKEADTD 384


```
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 03-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465567/50.
XX N-PSDB; AAS39532.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX disorders related to the colon including colon cancers and also for
XX testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID No 429; 562pp; English.
XX
XX The present invention relates to the isolation of novel human colon
XX associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic
XX sequences encoding for them. The sequences of the invention are useful
XX in the diagnosis, treatment, prevention and/or prognosis of disorders
XX of the colon including colon cancer, congenital abnormalities
XX (e.g. atresia and stenosis), bacterial and viral infections,
XX inflammatory bowel disease (IBD), neoplastic cell disorders,
XX colonic inflammation, diarrhoea and dysentery, malabsorption syndromes,
XX (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
XX The polynucleotides sequences of the invention can also be used in gene
XX therapy. AAU22468-AAU22701 represent the novel human colon associated
XX polypeptides of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 99 AA;
XX
XX Query Match 7.9%; Score 232; DB 22; Length 99;
XX Best Local Similarity 55.4%; Pred. No. 7.1e-16;
XX Matches 46; Conservative 15; Mismatches 22; Indels 0; Gaps 0;
XX
XX 1 MTITYTNKVNARLGSFSSLLCWGRGSIYKLLYGELVFIYLYSIRGLYVLSDDQL 60
XX 13 MTVSYTLKVAERRGFGSLLLRWGRGSIYKLLYKELFLFGALYAVLSYTYLLLTQERY 72
XX 61 LFEKALYCSYIQLIPISFVLG 83
XX 73 VYAOVARYCNRSADLIPLSFVLG 95
```

Search completed: July 10, 2003, 12:24:13
Job time : 35.021 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:34 ; Search time 12.6264 Seconds
(without alignments)
1283.984 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTNKANVARLGSPSSL.....EHASSYPYRDAGTKPVLVE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCITUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96.5	3.3	839	US-09-197-636-2	Sequence 2, Appli
2	96.5	3.3	839	US-09-197-636-4	Sequence 4, Appli
3	96.5	3.3	839	US-09-197-636-8	Sequence 8, Appli
4	96.5	3.3	839	US-09-235-451-34	Sequence 34, Appli
5	96.5	3.3	839	US-09-533-220A-2	Sequence 2, Appli
6	91.5	3.1	993	US-08-468-557-2	Sequence 2, Appli
7	89.5	3.0	838	US-09-235-451-2	Sequence 2, Appli
8	89.5	3.0	838	US-09-132-316-3	Sequence 3, Appli
9	89	3.0	1099	US-09-442-100-2	Sequence 2, Appli
10	88.5	3.0	997	US-08-233-540-1	Sequence 1, Appli
11	88.5	3.0	997	US-08-428-949A-1	Sequence 1, Appli
12	88.5	3.0	997	US-08-428-948A-1	Sequence 1, Appli
13	88.5	3.0	997	US-08-428-946-1	Sequence 1, Appli
14	88.5	3.0	997	PCT-US95-04656-1	Sequence 1, Appli
15	88.5	3.0	998	US-08-233-008A-6	Sequence 6, Appli
16	88.5	3.0	1021	US-08-233-008A-2	Sequence 2, Appli
17	88	3.0	688	US-09-016-000-2	Sequence 2, Appli
18	87	3.0	681	US-08-272-255-9	Sequence 6, Appli
19	87	3.0	681	US-08-272-255-9	Sequence 9, Appli
20	87	3.0	681	PCT-US95-08565-6	Sequence 6, Appli
21	87	3.0	681	PCT-US95-08565-9	Sequence 9, Appli
22	86.5	2.9	2304	US-09-324-867-4	Sequence 4, Appli
23	86	2.9	1956	US-08-843-417-10	Sequence 10, Appli
24	86	2.9	2319	US-08-212-133A-8	Sequence 8, Appli
25	86	2.9	2319	US-08-474-503-6	Sequence 6, Appli
26	86	2.9	2319	US-08-670-707A-6	Sequence 6, Appli
27	86	2.9	2319	US-09-037-601-6	Sequence 6, Appli

28 86 2.9 2319 4 US-09-315-179-6 Sequence 6, Appli
29 86 2.9 2319 4 US-09-523-656-28 Sequence 28, Appli
30 86 2.9 2319 5 PCT-US94-13200-6 Sequence 6, Appli
31 85.5 2.9 459 4 US-09-097-889-22 Sequence 22, Appli
32 85.5 2.9 684 4 US-09-134-078-62 Sequence 62, Appli
33 84.5 2.9 968 4 US-09-302-812-6 Sequence 6, Appli
34 84.5 2.9 968 4 US-09-511-477-6 Sequence 6, Appli
35 84.5 2.9 968 4 US-09-511-507-6 Sequence 6, Appli
36 84.5 2.9 1144 1 US-08-147-812-5 Sequence 5, Appli
37 84.5 2.9 1144 2 US-08-319-866-12 Sequence 12, Appli
38 84.5 2.9 1144 4 US-09-123-708-2 Sequence 2, Appli
39 84.5 2.9 1144 4 US-09-123-624-2 Sequence 2, Appli
40 84 2.9 681 3 US-08-984-268-5 Sequence 5, Appli
41 84 2.9 1984 3 US-08-836-325-10 Sequence 10, Appli
42 83.5 2.8 237 2 US-08-760-745-5 Sequence 5, Appli
43 83.5 2.8 554 4 US-09-134-001C-5109 Sequence 5109, Ap
44 83 2.8 626 2 US-08-617-801A-4 Sequence 4, Appli
45 83 2.8 857 1 US-07-717-331F-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-197-636-2
; Sequence 2, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-197-636-2

```
Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLYGFLVFIFLYSIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
DB 477 GEILSVLGG-----VYFFRGIQYEL-----QRRPSMKTLFVDSYSEM--LFFLOSFL 522
QY 86 VTLVSRWMSQYEN-----LPWPDRLMIQVSSFVEGKDEEGRLLRRLTLRYAILG 135
DB 523 MLATVVLVYFSLKLEYVASMVFSALGWTNMLY-----YTRGFQMG-----IYAVMI 569
QY 136 QVILIRISISVYKFRFTLHLVLGFWT-----HGEHKQLQKLGIPHNTPFVWPVWF 188
DB 570 EKMLRLDLC-----RPMFVIVFLFGSTAVVTLIEDGKNDLSPSESTSHR----- 615
QY 236 YQVTVVAVYSFPLACLIQROFLPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAYVILTILLNMLIALMGETVKNIAQESKNIKLQRAITITIDTEKSLKCMKRAFR 722
QY 285 MGWLKVAEQLINPGEDDDDF-----ETNWI-----IDRNLOVSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPPYTAASARSRHSFMGST-----FNISLKKEDELEL 375
DB 781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPEDAEV 829

QY 376 WSKKEADTK 385
DB 830 FKSPAASGEK 839

RESULT 2
US-09-197-636-4
; Sequence 4, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Prestia, Paul P
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30075
TELECOMMUNICATION INFORMATION:
TELEPHONE: 601-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 839 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-197-636-4

Query Match      3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

QY 26 GSIYKLYGFLVFIFLYSIRGLYRMVLSDDQQLFEKALYCDSYIQLIPISFVLGFY 85
DB 477 GEILSVLGG-----VYFFRGIQYEL-----QRRPSMKTLFVDSYSEM--LFFLOSFL 522
QY 86 VTLVSRWMSQYEN-----LPWPDRLMIQVSSFVEGKDEEGRLLRRLTLRYAILG 135
DB 523 MLATVVLVYFSLKLEYVASMVFSALGWTNMLY-----YTRGFQMG-----IYAVMI 569
QY 136 QVILIRISISVYKFRFTLHLVLGFWT-----HGEHKQLQKLGIPHNTPFVWPVWF 188
DB 570 EKMLRLDLC-----RPMFVIVFLFGSTAVVTLIEDGKNDLSPSESTSHR----- 615
QY 189 ANLSMKAYLGRIRDITVLLQSLMNEVCTLRQCQLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP---DSSYN---SLYSTCLELKFKTIGMDLEFTENYDKAVFII 662
QY 236 YQVTVVAVYSFPLACLIQROFLPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAYVILTILLNMLIALMGETVKNIAQESKNIKLQRAITITIDTEKSLKCMKRAFR 722
QY 285 MGWLKVAEQLINPGEDDDDF-----ETNWI-----IDRNLOVSLLS 321
DB 723 SG--KLLQVGYTPDGKDDYRCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPPYTAASARSRHSFMGST-----FNISLKKEDELEL 375
DB 781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPEDAEV 829

QY 376 WSKKEADTK 385
DB 830 FKSPAASGEK 839

RESULT 3
US-09-197-636-8
; Sequence 8, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

Fri Jul 11 11:45:18 2003

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/197.636
FILING DATE: 23-NOV-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 9805137.8
FILING DATE: 12-MAR-1998
APPLICATION NUMBER: UK 9815791.0
FILING DATE: 21-JUL-1998
APPLICATION NUMBER: UK 9819278.4
FILING DATE: 03-SEP-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30075
TELEPHONE: 601-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 839 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-197-636-8

Query Match 3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74; Indels 137; Gaps 19;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

26 GSIYKLLYGEFLVFIPLYSIRGLYRMVLSDDQQLLEKALYCDSYIQLIPISFVLGY 85
477 GEILSVLGG-----VYFFFRGIQYFL-----QRRPSMKTFLFVDSYSEM--LFFLQSLF 522
86 VTLVSRMWSQYEN-----LPWPDRLMIQVSSFVEGKDEBGLLRRLRLIRYAILG 135
523 MLATVLYFSLHKEYVASMVPSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
136 QVILRSISTSVKRPPTLHLVLAGEMT-----HGEHKQLQKLGPHNFTFWVWVWF 188
570 EKMLRLDLC-----RPMFVYVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
189 ANLSMKAYLGGIRDTVLLQSLMNEVCTLRTOCGQLYA-----YDWISIPLV 235
616 -----WRGPACRPP-----DSSYN---SLYSTCLELFKFTIGMGDLEFTENDYDFKAVFII 662
236 YTOVTVAVYSFPLACLIHQFLNPKNDYPGHE-----MDLVVPVFTILOFLFY 284
663 LLLAYVILTYILLNMLIALMGETVKNKIAQESKNWKLQRAITILDTEKSFCLKMKRAFR 722
285 MGWLKVAEQLINPFGEDDDF-----ETNMI-----IDRNLOVSLLS 321
723 SG--KLQVGYTPDGKDYRCWCFRVDENVNWTNTNNGIINEDPGNCEGVKRTLSFSLRS 780
322 --VDGMHONLPPMERDMYNEAAPQPYTAASARRSRHSFMGST-----FNISLKKEDLEL 375
781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPEDAEV 829
376 WSKKEADTDK 385
830 FKSPAASGEK 839

RESULT 5
US-09-533-220A-2
Sequence 2, Application US/09533220A
Patent No. 6406908
GENERAL INFORMATION:
APPLICANT: McIntyre, Peter
APPLICANT: James, Iain Fraser
TITLE OF INVENTION: Human Vanilloid Receptor
FILE REFERENCE: 4-30875A
CURRENT APPLICATION NUMBER: US/09/533,220A
CURRENT FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: UNITED KINGDOM 9907097.1
PRIOR FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent in Ver. 1.30
SEQ ID NO 2

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/197.636
FILING DATE: 23-NOV-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 9805137.8
FILING DATE: 12-MAR-1998
APPLICATION NUMBER: UK 9815791.0
FILING DATE: 21-JUL-1998
APPLICATION NUMBER: UK 9819278.4
FILING DATE: 03-SEP-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-30075
TELEPHONE: 601-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 839 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-197-636-8

Query Match 3.3%; Score 96.5; DB 4; Length 839;
Best Local Similarity 19.8%; Pred. No. 0.74; Indels 137; Gaps 19;
Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

26 GSIYKLLYGEFLVFIPLYSIRGLYRMVLSDDQQLLEKALYCDSYIQLIPISFVLGY 85
477 GEILSVLGG-----VYFFFRGIQYFL-----QRRPSMKTFLFVDSYSEM--LFFLQSLF 522
86 VTLVSRMWSQYEN-----LPWPDRLMIQVSSFVEGKDEBGLLRRLRLIRYAILG 135
523 MLATVLYFSLHKEYVASMVPSLALGWTNMLY-----YTRGFQOMG-----IYAVMI 569
136 QVILRSISTSVKRPPTLHLVLAGEMT-----HGEHKQLQKLGPHNFTFWVWVWF 188
570 EKMLRLDLC-----RPMFVYVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
189 ANLSMKAYLGGIRDTVLLQSLMNEVCTLRTOCGQLYA-----YDWISIPLV 235
616 -----WRGPACRPP-----DSSYN---SLYSTCLELFKFTIGMGDLEFTENDYDFKAVFII 662
236 YTOVTVAVYSFPLACLIHQFLNPKNDYPGHE-----MDLVVPVFTILOFLFY 284
663 LLLAYVILTYILLNMLIALMGETVKNKIAQESKNWKLQRAITILDTEKSFCLKMKRAFR 722
285 MGWLKVAEQLINPFGEDDDF-----ETNMI-----IDRNLOVSLLS 321
723 SG--KLQVGYTPDGKDYRCWCFRVDENVNWTNTNNGIINEDPGNCEGVKRTLSFSLRS 780
322 --VDGMHONLPPMERDMYNEAAPQPYTAASARRSRHSFMGST-----FNISLKKEDLEL 375
781 SRVSGRH-----WKNFALVPLLRASARDROSAQPEEVYLRQFSGSLKPEDAEV 829
376 WSKKEADTDK 385
830 FKSPAASGEK 839

RESULT 4
US-09-235-451-34
Sequence 34, Application US/09235451
GENERAL INFORMATION:
APPLICANT: Julius, David J.

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

```

NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-13A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 993 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-557-2

Query Match 3.1%; Score 91.5; DB 1; Length 993;
Best Local Similarity 19.6%; Pred. No. 3.2;
Matches 83; Conservative 53; Mismatches 134; Indels 153; Gaps 20;

QY 94 WSOYENLPWDRMIQVSSFEVGEKDEGRLLRRLIRYAILQGVILRLSISVYKRPPT 153
DB 588 WSYV-LPGPDQ---RADLYEGKQQLGWFQSSL-----TSVAARKKA 627
QY 154 LHLVLAGFTHGE--HKQLQKLG-----LPHNTFVVPVWVPA--NL 191
DB 628 PYKTVVHGFTLGEKGMKSLGNVHPDVVNGGQDQSKPEPYGADVLRW-WVADSNV 686
QY 192 SMKAYLG-----GRIDTV-----LLQS 209
DB 687 FTEAIGPSVLNAARDITSLKRLTLRFLGNVADFNPTDTSIPVNDMYVIDQYMLHLLQD 746
QY 210 LKNEVCTLRQCGQYADWISIPLVYVTVVAVYFFLACLGROFLNPKDYPGHE- 268
DB 747 LANKITELYQ-----YDFGVVRELLRTFYTRLSNFFYSIIKORLYCEKENDPKRSC 800
QY 269 -----MDLVVPVFT-ILQFLFYNGWLKVAEOLIN--PF-GEDDDDPFTWIDRNLOV 317
DB 801 QTALVEILDVIVRSFAPILPHL-----AEVQHPIYIPEKPSVPTGTWISTSSI-- 850
QY 318 SLLSVGDGHQNLPPMERDMYNEAAPQPPYTAASARSRRHSPMGS-----TF 364
DB 851 -----WKK--PGLBEAVESACAMRDSFLGSIPOKNAAEYKVIITVI 888
QY 365 NISLKEDLE-LMSKEADTKKESGYSSTIGFCGLQPKNYHLPLKDLTKLCKSNPL 423
DB 889 EPGLLEIEMLOSEETSTSQLNELMASESTLLAQEPREMTADVIELKGRFLIN---- 944
QY 424 LEG 426
DB 945 LEG 947

RESULT 7
US-09-235-451-2
Sequence 2, Application US/09235451
GENERAL INFORMATION:
APPLICANT: Julius, David J.
APPLICANT: Caterina, Michael J.
APPLICANT: Brake, Anthony J.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 9076/084CIP
CURRENT APPLICATION NUMBER: US/09/235,451
CURRENT FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/072,151
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: 08/915,461
PRIOR FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 838
TYPE: PRT

```

[illegible]

RESULT 9
 US-09-442-100-2
 ; Sequence 2, Application US/09442100
 ; Patent No. 6359193
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Wang, Weiyl
 ; APPLICANT: Zhang, Sheng
 ; APPLICANT: Yu, Nan
 ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
 ; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM: disk
 ; MEDIUM TYPE: IBM PC compatible
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/09/442,100
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/411,111
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hirstock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 6523-003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1099 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein

ORGANISM: R. rattus
US-09-235-451-2

Query Match 3.0%; Score 89.5; DB 4; Length 838;
Best Local Similarity 18.5%; Pred. No. 3.9;
Matches 87; Conservative 61; Mismatches 157; Indels 165; Gaps 22;

24 WRGSYKLLYGEFLV-----IP-----LYYSIRGL--YRM-----VLSSDQ 59
426 WDRFVKRIFYFNFFVCLYMIIFTAAAYRYPVEGLPYKLNKTVGDYFRVTEILSVSGG 485
60 LLF-----EKALYCDYSYIQLIPISVFLGFTVTLVSRWMSQYEN-----99
486 VYFFFRGIGYFLQRPSLKLSFVDSYSEI--LPFVQSLPMLVSVLYFSQRKEYYASVMP 543
100 ----LWPDRLMIQVSVFVEGKDBEGRLLRFTLIRYAILQVLTIRISIVYKRPPTLHH 156
544 SLANGTWMLY-----YTRGFQOMG-----IYAVMIKMLRLDLC--RMPFVYL 585
157 LVLAGFWT-----HGEHKQLQKGLPHNTFWPWFVWFANLSMKAYLGRIRDVTLVLOS 209
586 VPLFGFSTAVVTLLIEDGKNLSPMESTPH-----KCRGSACKRG 624
210 LMNEVCTRLTQCGOLYA-----YDMISPLVYQVTVVAVYSFFELACLIGRQ 256
625 --NSYSLYSTCLELKFKTIGMDLEFTENYDFKAVFILLAVIITYILLNMLIALM 682
257 FLNPNKYPGHE-----MDLVVPVFTILOFLFVGMGLKVAQLNPFGEDDDDF 305
683 GETVKNIAQESKNIMKWLQRAITLIDTEKSEFLKCMKAFRSG--KLQVGFPTDGDYRW 740
306 ----ETNWI-----IDRNQVSLLS--VDGMHQNLPPMERDMYWNE 340
741 CERUDEVNWTWNTNVGIINEDPCNCEGVKRTLSFLSRGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSFMGSFNI-----SLKKEDLSLKEEADTK 385
790 FALVPLLMQASTRD-RHATQOEVEOLKHYTGSLKPEDAEVFXDSMVPGEK 838

RESULT 8
US-09-132-316-3
Sequence 3, Application US/09132316B
Patent No. 6444440
GENERAL INFORMATION:
APPLICANT: Young, Steven M.
TITLE OF INVENTION: Vanilloid Receptor-2
FILE REFERENCE: 1488.1110000
CURRENT APPLICATION NUMBER: US/09/132,316B
CURRENT FILING DATE: 1998-08-11
EARLIER APPLICATION NUMBER: US 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 838
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-132-316-3

Query Match 3.0%; Score 89.5; DB 4; Length 838;
Best Local Similarity 18.5%; Pred. No. 3.9;
Matches 87; Conservative 61; Mismatches 157; Indels 165; Gaps 22;

24 WRGSYKLLYGEFLV-----IP-----LYYSIRGL--YRM-----VLSSDQ 59
426 WDRFVKRIFYFNFFVCLYMIIFTAAAYRYPVEGLPYKLNKTVGDYFRVTEILSVSGG 485
60 LLF-----EKALYCDYSYIQLIPISVFLGFTVTLVSRWMSQYEN-----99
486 VYFFFRGIGYFLQRPSLKLSFVDSYSEI--LPFVQSLPMLVSVLYFSQRKEYYASVMP 543
100 ----LWPDRLMIQVSVFVEGKDBEGRLLRFTLIRYAILQVLTIRISIVYKRPPTLHH 156
544 SLANGTWMLY-----YTRGFQOMG-----IYAVMIKMLRLDLC--RMPFVYL 585
157 LVLAGFWT-----HGEHKQLQKGLPHNTFWPWFVWFANLSMKAYLGRIRDVTLVLOS 209
586 VPLFGFSTAVVTLLIEDGKNLSPMESTPH-----KCRGSACKRG 624
210 LMNEVCTRLTQCGOLYA-----YDMISPLVYQVTVVAVYSFFELACLIGRQ 256
625 --NSYSLYSTCLELKFKTIGMDLEFTENYDFKAVFILLAVIITYILLNMLIALM 682
257 FLNPNKYPGHE-----MDLVVPVFTILOFLFVGMGLKVAQLNPFGEDDDDF 305
683 GETVKNIAQESKNIMKWLQRAITLIDTEKSEFLKCMKAFRSG--KLQVGFPTDGDYRW 740
306 ----ETNWI-----IDRNQVSLLS--VDGMHQNLPPMERDMYWNE 340
741 CERUDEVNWTWNTNVGIINEDPCNCEGVKRTLSFLSRGRVSGRN-----WKN 789
341 AAPQPPYTAASARRSRHSFMGSFNI-----SLKKEDLSLKEEADTK 385
790 FALVPLLMQASTRD-RHATQOEVEOLKHYTGSLKPEDAEVFXDSMVPGEK 838

us-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

NAME: Robinson, Joseph R.
 REGISTRATION NUMBER: 33,448
 REFERENCE/DOCKET NUMBER: 0646/1B017-USA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7783
 TELEFAX: 212-753-6327
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 997 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-428-948A-1

Query Match 3.0%; Score 88.5; DB 1; Length 997;
 Best Local Similarity 19.0%; Pred. No. 6.6;
 Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSOQQLFEKALYCDSYIQLIPISFVLGFTVTLVSRWWSQYENLPWPDRLMIQVS-SF 113
 DB 300 SODKQLF-----DNYV-----ILGNYTT-----LMFNISRAY 326
 QY 114 VEGKDEGRLLRRLIRYAILQVILRSISTSVYKRPFTLHLVLGPMTHGEHKQLOK 173
 DB 327 VLEKDPQK-----AQLKQMYLLMT-----KHLDDQGFVKGSAVLTTHH 365
 QY 174 LGLPHNTFWPWFVWFWANLSKAVLGGRIKDTVLLOSL-----MNEVCTL 217
 DB 366 WGYSSRWYISTLLMSDALKEANLQTVYDLSLLWYSREFKSSPDMKVSADSDLDYFNTL 425
 QY 218 RTQCGQLYADWISIPLVYTVQVTVAVVSPFLACLI-----GRQFLNPN-----KDY 264
 DB 426 SRQHALLLLE-----PDQKRLNLTFTSHYITGALTQVPPGKGDLRDPGTAWRHEGNY 481
 QY 265 PGHENDLVVPVFTILOFLFYM-----GWLKVAEQLINPFGEDDDDDFTETNMLDR 313
 DB 482 PGYSF-----PAFKNASQLIYLLRDTFVSFGESGNNLKAMVSA-----WYS- 525
 QY 314 NLOVSLLSVDGMH-----QNLPPMERDMYW-----NEAAPQPP-----YTAASARSRHS--- 358
 DB 526 NPEVG-LPLAGRHFPNSPLSKVAQGYWLAWSAKSSPDKTLASIYLAISDKTQNESTAI 584
 QY 359 -----FMGSTFNI-----SLKEDLELWSKEADTKKESGYSS-- 392
 DB 585 FGETITPASLPQGFYAFNGGAFGIHRWQDKMVTLKAYNTNWSSEIYNKNRYCRYOSH 644
 QY 393 -----TIGCFLL--GLQPKNY-----HLPLKDL 412
 DB 645 VAQIVNSGSQLSQGYQOEGWDMWNRMQGATTIHLPLKDL 682

RESULT 13

US-08-428-946-1
 ; Sequence 1, Application US/08428946
 ; Patent No. 5855883
 ; GENERAL INFORMATION:
 ; APPLICANT: Kiran M. Khandke, John Gatto, Ursula Eul
 ; TITLE OF INVENTION: Method of Disinsection of Vitreous Body from
 ; TITLE OF INVENTION: Neural Retina of the Eye
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby P.C.
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

RESULT 12

US-08-428-948A-1
 ; Sequence 1, Application US/08428948A
 ; Patent No. 5741692
 ; GENERAL INFORMATION:
 ; APPLICANT: Khandke, Kiran M.
 ; TITLE OF INVENTION: Chondroitinase II
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby P.C.
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/428,948A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

US-09-622-964-29.ra1

Fri Jul 11 11:45:18 2003

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04656
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gordon, Alan M.
 REGISTRATION NUMBER: 30,637
 REFERENCE/DOCKET NUMBER: 32,390-00/PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-831-3244
 TELEFAX: 201-831-3305
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 997 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04656-1

Query Match 3.0%; Score 88.5; DB 5; Length 997;
 Best Local Similarity 19.0%; Pred. No. 6.6;
 Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSDQQLLFKALYCDYSYIQLIPISFVLGVYVTLVSRWWSQYENLPWPDRLMIQV-SF 113
 DB 300 SQDKQLF-----DNYV-----ILGNYYT-----LMFNISRAY 326
 QY 114 VEGKDEGRLLRRLIRYAILGOVLILRSISTSVYKRFPTLHLVLGFWTHGEHKQLOK 173
 DB 327 VLEKDPQK-----AQKQWYLLMT-----KHLDDQGFVKGSALVYTHH 365
 QY 174 LGLPHNTFWPWWFANLKMAYLGGRIQRTVLQSL-----MNEVCTL 217
 DB 366 WGYSSRWYIITLMSDALKEANLQTVDSLWYSREFKSSPDMKVSADSSLDYFNTL 425
 QY 218 RTQCQLVAYDWISPIVYTVQVTVAVYSPFLACLI-----GRQFLNPN-----KDY 264
 DB 426 SRQLALLLLE-----PDQKRLNLTFSHYITGALTQVPPGGKGLRPGDTAWRHGNY 481
 QY 265 PGHEMDLVVPVFTILOFLFYM-----GWLKVAEQLINPFGEDDDDDFTNWIIDR 313
 DB 482 PGYSF-----PAFKNASQLIYLLRDTFPFVGSGWNNLKAMVSA-----WYS- 525
 QY 314 NLQVSLLSVDGMH-----QNLPPMERDMY-----NEAAPQPP-----YTAASARSRRHS--- 358
 DB 526 NPEVG-LPLAGRHPNPSPLKSAQGYWYLAWSAKSSPDKTSLASIYLAISDKTQNESTAI 584
 QY 359 -----FMGSTFNI-----SLKKEDELSKEEADTDKESGYSS-- 392
 DB 585 FGTTITPASPQGFYAFNGGAFGIHRWQDKMVTLKAYNTNWSSEIYNKDNRYGRYQSHG 644
 QY 393 -----TIGCFPL-----GLQPKNY-----HLPLKDL 412
 DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 15

US-08-233-008A-6
 ; Sequence 6, Application US/08233008A
 ; Patent No. 5578480
 ; GENERAL INFORMATION:
 ; APPLICANT: Khandke, Kiran M.
 ; TITLE OF INVENTION: Methods For The Isolation And
 ; TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
 ; TITLE OF INVENTION: I and II Enzymes From P. vulgaris
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: American Cyanamid Company
 ; STREET: One Cyanamid Plaza
 ; CITY: Wayne

APPLICATION NUMBER: US/08/428,946
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Robinson, Joseph R.
 REGISTRATION NUMBER: 33,448
 REFERENCE/DOCKET NUMBER: 0846/1B017-US3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7783
 TELEFAX: 212-753-6237
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 997 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-428-946-1

Query Match 3.0%; Score 88.5; DB 2; Length 997;
 Best Local Similarity 19.0%; Pred. No. 6.6;
 Matches 87; Conservative 62; Mismatches 134; Indels 175; Gaps 23;

QY 55 SSDQQLLFKALYCDYSYIQLIPISFVLGVYVTLVSRWWSQYENLPWPDRLMIQV-SF 113
 DB 300 SQDKQLF-----DNYV-----ILGNYYT-----LMFNISRAY 326
 QY 114 VEGKDEGRLLRRLIRYAILGOVLILRSISTSVYKRFPTLHLVLGFWTHGEHKQLOK 173
 DB 327 VLEKDPQK-----AQKQWYLLMT-----KHLDDQGFVKGSALVYTHH 365
 QY 174 LGLPHNTFWPWWFANLKMAYLGGRIQRTVLQSL-----MNEVCTL 217
 DB 366 WGYSSRWYIITLMSDALKEANLQTVDSLWYSREFKSSPDMKVSADSSLDYFNTL 425
 QY 218 RTQCQLVAYDWISPIVYTVQVTVAVYSPFLACLI-----GRQFLNPN-----KDY 264
 DB 426 SRQLALLLLE-----PDQKRLNLTFSHYITGALTQVPPGGKGLRPGDTAWRHGNY 481
 QY 265 PGHEMDLVVPVFTILOFLFYM-----GWLKVAEQLINPFGEDDDDDFTNWIIDR 313
 DB 482 PGYSF-----PAFKNASQLIYLLRDTFPFVGSGWNNLKAMVSA-----WYS- 525
 QY 314 NLQVSLLSVDGMH-----QNLPPMERDMY-----NEAAPQPP-----YTAASARSRRHS--- 358
 DB 526 NPEVG-LPLAGRHPNPSPLKSAQGYWYLAWSAKSSPDKTSLASIYLAISDKTQNESTAI 584
 QY 359 -----FMGSTFNI-----SLKKEDELSKEEADTDKESGYSS-- 392
 DB 585 FGTTITPASPQGFYAFNGGAFGIHRWQDKMVTLKAYNTNWSSEIYNKDNRYGRYQSHG 644
 QY 393 -----TIGCFPL-----GLQPKNY-----HLPLKDL 412
 DB 645 VAQIVNSGSQLSQGYQOEGWDMNRMOGATTIHLPLKDL 682

RESULT 14

PCT-US95-04656-1
 ; Sequence 1, Application PC/TUS9504656
 ; GENERAL INFORMATION:
 ; APPLICANT: Khandke, Kiran M.
 ; TITLE OF INVENTION: A Novel Protein Designated
 ; TITLE OF INVENTION: Chondroitinase II and its Use with a Protein Designated
 ; TITLE OF INVENTION: Chondroitinase I to Achieve Complete Vitreal Disinsertion
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: American Cyanamid Company
 ; STREET: One Cyanamid Plaza
 ; CITY: Wayne
 ; STATE: New Jersey
 ; COUNTRY: U.S.A.
 ; ZIP: 07470-8426
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

Search completed: July 10, 2003, 12:27:57
Job time : 16.6264 secs

	Best, Local Similarity	53.8t;	Pred. No. 1.3e-52;	Mismatches 31;	Indels 18;	Gaps 4
QY	179	NTFWPTWFWANLSKAVILGRIEDT	VLLQSLMNEVCTLR	TCGQQLYXDWDISIP	LVYVQ	238
Db	20	NKYWYPCVWF	TNLAQAQRDRGRI	RDIDTALCLLEELNKYRAKCSMLFHYDWDISIP	LVYVQ	79
QY	239	VVTVAVYSFF	FLACLRIGROFLNPKD	-----YPGHE-----	-MDLVVPVETILQFLP	283
Db	80	VVTVAVYSFF	ALSLVGRQFVEPEAGA	KPKQLKPGQEPAPALGDPDMYVPLTLLQ	QFFF	139
QY	284	YMGWLKVAEQ	LINPFGE	DDDDFETNWTIIDRLNQLSVLLSDVGMHQNLPPMERD	MYWNEAP	343
Db	140	YAGWLKVAEQ	LINPFGE	DDDDFETNQLIDRLNQLSVLLSDVEMYNLPPAEKQDYW	DEDOQ	1999
QY	344	QPPYT--ASAR	RRRRHSFMGSTFTN	SLKDELDLMSKEADTD	DKKESGYST--	1GCFGLG 400
Db	200	QPPYTVATAAES	LRFSFLGSTFTNRMS	DDPQSLQVENS	PGSGRPPAAATPL	GLRGFLGV 259

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Milasincic, Debra J.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-746-783-4

Query Match 15.1%; Score 444; DB 9; Length 261;
Best Local Similarity 41.3%; Pred. No. 5.8e-34;
Matches 109; Conservative 34; Mismatches 71; Indels 50; Gaps 10;

QY 325 MHQNLPPMERDMYNEAAPQPPYTAASARRSHSGTFTNLSLKEDLELWSKEADTD 384
DB 1 MHQDLPRMEPDYMNKPEPPYTAASAQFRASPMGSTFTNLSKEMEFPQNGE----- 56
QY 385 KXESGVYSTIGTCFLGQPKNYHLPLKDLTKLCSKNPLL--EGCKD-----ANQKNQK 437
DB 57 DEEDAHAGIIGRFLGQSHDHPHPRANSRTKLLMPKRESLLHEGLPKNHKAQKQNVGOE 116
QY 438 D--VWKFGLDLFLKCVPRFKRGSHCGPOAPSS-----HPTQSPSS--SDTG----- 482
DB 117 DNKAWKLKAVDAFKSAPLYQRPQYSGAPQTPLSPTPMFEPFSPAPSKLHSVTGIDTKK 176
QY 483 -----DGPSTDYQEIChMKKKTVEFNL-NIPESPTHLQQRRLQDM 522
DB 177 SLKTVSSGAKKSPELLSESDGALMEHPVSVQVRKTVFNLDMPEIPENHLKE-PLBQS 235
QY 523 STNIQALMKHAEBSY---PYRDEA 543
DB 236 PTNIHTTIKOHMDPYALENDEA 259

RESULT 4
US-09-764-872-429
Sequence 429, Application US/09764872
Publication No. US20030050231A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA125
CURRENT APPLICATION NUMBER: US/09/764,872
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 957
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 429
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-872-429

Query Match 7.9%; Score 232; DB 9; Length 99;
Best Local Similarity 55.4%; Pred. No. 2.1e-14;
Matches 46; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 1 MTITTYNKVANARLGSFSLLCWFGSIYKLLYGEFLVFIYYSIGLFRWLSSQQL 60
DB 13 MTVSYTLKVAEARFGFGSGLLRWEGSIYKLLYKEFLFGALYAVLSITVRLLTQORY 72
QY 61 LFEKALYCDSYIQLIPISFVLG 83

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Milasincic, Debra J.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-746-783-4

Query Match 18.1%; Score 534; DB 10; Length 251;
Best Local Similarity 55.9%; Pred. No. 1.5e-42;
Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LVAYDWISIPLVYQVTVAVYSFPLACLIGRQFLNPNKD-----YFGE----- 268
DB 2 LFHYDWISIPLVYQVTVAVYSFPLSLVGRQFVEPEAGAARKPKLLKPGQEPAPALGD 61
QY 269 MDLVVPVFTILQFLPMGWLKVAEQIINPFGEDEDDDDFTNWIIDNLOVSLLSVDGWHQ 328
DB 62 PDWYVPLTLTLLQFFYAGWLKVAEQIINPFGEDEDDDDFTNLOIDRNLOVSLLSVDENIQ 121
QY 329 LPMEEDMYNEAAPQPPYTAASARRSHSGTFTNLSLKEDLELWSKEADTDKCE 387
DB 122 LPPEAKQDQWDEQPPYTVATAESLRPFLSGFTNLRMSDDPEQSLQVSEASPGSGRP 181
QY 388 SGYSST--IGCFGL 400
DB 182 APAAQTPLLRFLGV 196

RESULT 3
US-09-746-783-4
Sequence 4, Application US/09746783
Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Hoses, Steven H.
Fechtel, Kim
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 231
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
Version #1 30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/768,826
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Shi et al.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-09-768-826-35

Query Match 18.1%; Score 534; DB 10; Length 251;
Best Local Similarity 55.9%; Pred. No. 1.5e-42;
Matches 109; Conservative 24; Mismatches 44; Indels 18; Gaps 4;

QY 224 LVAYDWISIPLVYQVTVAVYSFPLACLIGRQFLNPNKD-----YFGE----- 268
DB 2 LFHYDWISIPLVYQVTVAVYSFPLSLVGRQFVEPEAGAARKPKLLKPGQEPAPALGD 61
QY 269 MDLVVPVFTILQFLPMGWLKVAEQIINPFGEDEDDDDFTNWIIDNLOVSLLSVDGWHQ 328
DB 62 PDWYVPLTLTLLQFFYAGWLKVAEQIINPFGEDEDDDDFTNLOIDRNLOVSLLSVDENIQ 121
QY 329 LPMEEDMYNEAAPQPPYTAASARRSHSGTFTNLSLKEDLELWSKEADTDKCE 387
DB 122 LPPEAKQDQWDEQPPYTVATAESLRPFLSGFTNLRMSDDPEQSLQVSEASPGSGRP 181
QY 388 SGYSST--IGCFGL 400
DB 182 APAAQTPLLRFLGV 196

RESULT 3
US-09-746-783-4
Sequence 4, Application US/09746783
Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Hoses, Steven H.
Fechtel, Kim
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 231
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
Version #1 30

```


	Best Local Similarity	19.8%; Pred.No. 4.3;	Mismatches 85; Conservative	54; Mismatches 154; Indels 137; Gaps	19;
Qy	26	GSIYKLLYGEBFVLFIPLYYISIRGLRMYLSSDQQLLFKALYCDSYIOLIPISFVIGFY	85		
Dd	477	GEILSVLG-----VIFFFRGIQYFL-----QRPSMKTILFVDYSYM--LFFLOSLF	522		
Qy	86	VTLVSRWMSQVEN-----LPWPRLMIQVSFVEGKDEEGRLLRRTLIRVAAILG	135		
Dd	523	MLATVVLYFSHUKEYVASMVFSIALGWTNLY-----YTRGFQQMG-----IYAVMI	569		
Qy	136	OVLILRESISTSVYKRPTLIHLVLACGPMT-----HGBHKQLQKLGLPHNTFWVPWWVF	188		
Dd	570	EKWILLRDLC-----RFMFVVVFLFGSTAAVTLIBDGKNDLSPESESTR-----	615		
Qy	189	ANLSMKAYLGGRIIDTVLLQSLENEVCTIRTCCQLYA-----YDWISIPLV	235		
Dd	616	-----WRGPA CRPP-----DSSYN-----SLSYSTCLELFTTIGMDLEFTENYDFKAVFII	662		
Qy	236	YTQVTVAVVSPFLACLIGRQLPNPKNDYPGHE-----MDLVVPVFITILQFLFY	284		
Dd	663	LILAIVILTYILLNLNLMIALMGETVNKIAQESKNWKLOQRAITILDTEKSFLCKMRKAER	722		
Qy	285	MGWLKVAEQLINPFBDDDP-----ETNI-----IDRNMQVSLLS	321		
Dd	723	SG- KLLQVGTYPDGKDYRCWFVRDEVNWTWNVTNVGIINEDPGNCCEGVKRTLSFSLRS	780		
Qy	322	--VDGHQHONLPPWERDMYNEAAPPYTAASARSRRHSFMGST----FNISLKEDLEL	375		
Dd	781	Srvsgrh-----WKNFALVPLLREASARDQSAQPBEVYLQFSGSLKPDAEV	829		
Qy	376	WSKERADTDK	385		
Dd	830	FKSPAASGER	839		

RESULT 9
US-09-824-258-2
Sequence 2, Application US/09824258
Patent No. US20010047090A1
GENERAL INFORMATION:
APPLICANT: DUCKWORTH, DAVID MALCOLM
APPLICANT: HAYES, PHILIP DAVID
APPLICANT: MEADOWS, HELEN JANE
APPLICANT: DAVIS, JOHN BERESFORD
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP-30075-D1
CURRENT APPLICATION NUMBER: US/09/824,258
CURRENT FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: UK 9805137.8
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: UK 9815791.0
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: UK 9819278.4
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: US 09/197,636
PRIOR FILING DATE: 1998-11-23
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 839
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-824-258-2.

	Query Match	3.3%; Score 96.5; DB 10; Length 839;			
	Best Local Similarity	19.8%; Pred.No. 4.3;	Mismatches 85; Conservative	54; Mismatches 154; Indels 137; Gaps	19;
Qy	26	GSIYKLLYGEBFVLFIPLYYISIRGLRMYLSSDQQLLFKALYCDSYIOLIPISFVIGFY	85		
Dd	477	GEILSVLG-----VIFFFRGIQYFL-----QRPSMKTILFVDYSYM--LFFLOSLF	522		

```
QY 86 VTLVSRWMSQYEN-----LPWDRMLQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLKKEYVASMVPSLALGWTNNLY-----YTRGFQOMG-----IYAVMI 569
QY 136 QVLLRSISTSVYKRPPTLHLVLAGFMT-----HGEHKOLQKLGPHNTFWVWVWF 188
DB 570 EXMILRDLC-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN---SLYSTCLELFPKFTIGMGDLEFTENYDFKAVFII 662
QY 236 YQOVTVAVYSPFLACLIGRQFLNPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAYVILTYILLNNLIALMGETVKNIAQESKNIWKLQRAITILDTEKSFCLKMKRKA 722
QY 285 MGWLKVAEQLINPFGEDDDDF-----ETNWI-----IDRLQVLSLS 321
DB 723 SG--KLLQVGYTPDGKDDYRWCFRVDENVNTWNTNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPYTAASARRHRHSMFGST-----FNISLKEDLEL 375
DB 781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEVYLRFQSGSLKPEDAEV 829
QY 376 WSKBEADTDK 385
DB 830 FKSPAASGEK 839
```

RESULT 10

```
US-09-824-258-4
; Sequence 4, Application US/09824258
; Patent No. US20010047090A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID MALCOLM
; APPLICANT: HAYES, PHILIP DAVID
; APPLICANT: MEADOWS, HELEN JANE
; APPLICANT: DAVIS, JOHN BERESFORD
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30075-D1
; CURRENT APPLICATION NUMBER: US/09/824,258
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: UK 9805137.8
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: UK 9815791.0
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: UK 9819278.4
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 09/197,636
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 839
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (144)(194)(198)
US-09-824-258-4
```

Query Match 3.3%; Score 96.5; DB 10; Length 839;

Best Local Similarity 19.8%; Pred. No. 4.3;

Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

```
QY 26 GSIYKLLYGEFLVFIPLYYSIRGLYRWLVSSDQQLLFEKALYCDVYIQLIPISFVLGY 85
DB 477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLEVDYSYSEM--LFFQLSLF 522
QY 86 VTLVSRWMSQYEN-----LPWDRMLQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLKKEYVASMVPSLALGWTNNLY-----YTRGFQOMG-----IYAVMI 569
```

```
QY 136 QVLLRSISTSVYKRPPTLHLVLAGFMT-----HGEHKOLQKLGPHNTFWVWVWF 188
DB 570 EXMILRDLC-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN---SLYSTCLELFPKFTIGMGDLEFTENYDFKAVFII 662
QY 236 YQOVTVAVYSPFLACLIGRQFLNPNKDYPCHE-----MDLVVPVFTILOFLFY 284
DB 663 LLLAYVILTYILLNNLIALMGETVKNIAQESKNIWKLQRAITILDTEKSFCLKMKRKA 722
QY 285 MGWLKVAEQLINPFGEDDDDF-----ETNWI-----IDRLQVLSLS 321
DB 723 SG--KLLQVGYTPDGKDDYRWCFRVDENVNTWNTNGIINEDPGNCEGVKRTLSFSLRS 780
QY 322 --VDGMHQNLPMPERDMYNEAAPQPYTAASARRHRHSMFGST-----FNISLKEDLEL 375
DB 781 SRVSGRH-----WKNFALVLLREASARDRQSAQPEVYLRFQSGSLKPEDAEV 829
QY 376 WSKBEADTDK 385
DB 830 FKSPAASGEK 839
```

RESULT 11

```
US-09-824-258-8
; Sequence 8, Application US/09824258
; Patent No. US20010047090A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID MALCOLM
; APPLICANT: HAYES, PHILIP DAVID
; APPLICANT: MEADOWS, HELEN JANE
; APPLICANT: DAVIS, JOHN BERESFORD
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30075-D1
; CURRENT APPLICATION NUMBER: US/09/824,258
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: UK 9805137.8
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: UK 9815791.0
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: UK 9819278.4
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 09/197,636
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 839
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-824-258-8
```

Query Match 3.3%; Score 96.5; DB 10; Length 839;

Best Local Similarity 19.8%; Pred. No. 4.3;

Matches 85; Conservative 54; Mismatches 154; Indels 137; Gaps 19;

```
QY 26 GSIYKLLYGEFLVFIPLYYSIRGLYRWLVSSDQQLLFEKALYCDVYIQLIPISFVLGY 85
DB 477 GEILSVLGG-----VYFFPRGIQYFL-----QRRPSMKTLEVDYSYSEM--LFFQLSLF 522
QY 86 VTLVSRWMSQYEN-----LPWDRMLQVSSFVEGKDEBGRLLRRTLIRYAILG 135
DB 523 MLATVVLVYFSLKKEYVASMVPSLALGWTNNLY-----YTRGFQOMG-----IYAVMI 569
QY 136 QVLLRSISTSVYKRPPTLHLVLAGFMT-----HGEHKOLQKLGPHNTFWVWVWF 188
DB 570 EXMILRDLC-----RPMFVYIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR-----615
QY 189 ANLSMKAYLGGRIRODTVLLQSLMNEVCTLRTOCGOLYA-----YDWISIPLV 235
DB 616 -----WRGPACRPP-----DSSYN---SLYSTCLELFPKFTIGMGDLEFTENYDFKAVFII 662
```

236 YQVVTAVYVFFLAELIGROFLNPKDYPGHE-----MDLVVPVFTILQIFLY 284
Db
663 LLLAVILTYILLNMLIAGETVKNIAQESKNWIKLQRAITILDTKSLKCKRAFR 722
Qy
285 MGWLKVAEQLNPFGEEDDDF-----ETNWI-----IDRLQVSLLS 321
Db
723 SG--KLLQGVTPDKDYRCFRVDEVNWTNTVNGIINEDPGNCEGVKRTLSFSLRS 780
Qy
322 --VDGMHQLPMPERDMYNEAAPQPYTAASARRHSFWGST-----FNISLKEDLEL 375
Db
781 SRVSGRH-----WKNFALVPLLRASARDQSAQPEEVYLRQFSGSLKPEDAEV 829
Qy
376 WSKKEADTDK 385
Db
830 FKSPAASGEK 839
RESULT 12
US-09-832-292-29
; Sequence 29, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 601-1-098CIP
; CURRENT APPLICATION NUMBER: US/09/832,292
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 09/632,131
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 29
; LENGTH: 1863
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-832-292-29
Query Match 3.3%; Score 96.5; DB 9; Length 1863;
Best Local Similarity 20.2%; Pred. No. 14;
Matches 129; Conservative 76; Mismatches 179; Indels 255; Gaps 39;
Qy 35 EFLVFTFLY-YSIRGLVRVLSDDQLLFEKALYCDYSIQ-----LIPISFVLGYVTVLV 89
Db 886 EWIVIAFTYALEKV-REVFMSEAGKISQIKWFSDFYVSDTIALISFFVGFGLRF- 943
Qy 90 VSRW--WSQYEN-----LPWDRMIQVSSFVEGKDEBGRLLRRLTYAI 133
Db 944 GAKWYINAYDNHVFVAGRLIYCLNIIFWYVRL-----DFLAVNQAGPYV----- 990
Qy 134 LGQVLIURSTSVYKRPPTLHLVLVAGPMTHGEHKOLKGLPHNTFWVP-----WVWF 188
Db 991 ---MMIGWVANMFY---IVVIMALVLSF-----GVPRKAILPHEPWSLA 1033
Qy 189 ANLSMKAY--LGGRIKDTVLQSLMNEVC-----TLRTQCG-----QLYAYDWI 230
Db 1034 KDIVFHPYWMIFGEV-----YAYEIDVCANDSTLPTICPGTWLTPFLQAVLVFQYII 1087
Qy 231 SIPL-----VYTVQVTVV-----VYSFPLACLIGRQFLNPKDYPGHEMDLVVPVF 276
Db 1088 MVNLLIAFFNNVYLQVRAISIVWKYQRYHEIMAY-----HEKPLPPL 1132
Qy 277 TILQ-----FLVGMGLKV-----AEQLINP-FGEEDDDDPET- 307
Db 1133 IILSHIVSLFCVCKRRKDKTSDGPKLFLTEEDOKLHDEQCVMFYFEKDDKNSG 1192
Qy 308 -----NMIIIDRLQVSLLSVDGMHQLNLPMPERDMYNEA 341
Db 1193 SEERIRVTFERVEQMSIQIKEVGDRVNY-IKESLQ-SLDSQIGHQLDLSALTVDTLTLT 1250
Qy 342 APQPPYTAASARRHSFWGSTFNISLKK-----EDL-----ELMSKEADTDKESGYS 391

1251 AQK-----ASEASKVNEI--TRELSISKHLAQLNIDDDVPVRLWKKPSAVNTLSSS--- 1300
Qy 392 STTGCEFLGLQPKNYHLPLKDLTKLCSKNPLLEGCKDANKQKQD-VVKFGLDPLKC 450
Db 1301 -----LPQGDRE-----SNNPFL---C-NIFMKDEKDPQTNLFQGD-LPV 1335
Qy 451 VPRFK-----RRSHGCPQAPSS-----HPTEQ-----SAPSSSDTGDG 484
Db 1336 IPORKEFNIPKESGCCALFPFSAVSPPELQRHGVEMLKI FNKNQKLGSPNPSPP----- 1391
Qy 485 PSTDYQIECHMKKKTVEFNINIPESPT--EHLQORRLDQ 521
Db 1392 -----HMSPPPKFSVSTPSQSCSKHLESTTKDQ 1421
RESULT 13
US-09-764-367A-7
; Sequence 7, Application US/09764367A
; Patent No. US20020072101A1
; GENERAL INFORMATION:
; APPLICANT: Gaughan, Glen
; APPLICANT: Ramanathan, Chandra
; TITLE OF INVENTION: NOVEL HUMAN NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING CAT
; TITLE OF INVENTION: CHANNELS
; FILE REFERENCE: 5624.252.999
; CURRENT APPLICATION NUMBER: US/09/764,367A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/177,554
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 7
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-367A-7
Query Match 3.2%; Score 95; DB 10; Length 829;
Best Local Similarity 19.6%; Pred. No. 5.9; Mismatches 53; Indels 132; Gaps 17;
Matches 83; Conservative 53;
Qy 26 GSIYKLLYGEFLVFIIFYYSIR---GLYRMVLSDDQLLFEKALYCDYSIQIIPISFVL 82
Db 476 GEILSVLGVYFFFGIYQYFLQRRPSMKTLPFDSYSEMLFFLQSLFMLATVVLASNVFSL 535
Qy 83 GFVTLVSRMWSQYENLPWDRMIQVSSFVEGKDEBGRLLRRLTYAILQGVLLRS 142
Db 536 A-----LGWTNMLY-----YTRGFQOMG-----IYAVMIERKMILRD 566
Qy 143 ISTSVYKRPPTLHLVLVAGPM-----HGEHKOLKGLPHNTFWVPWVWFAVLSMKA 195
Db 567 LC-----RFMFYIIVFLFGFSTAVVTLIEDGKNDLSPSESTSHR----- 605
Qy 196 YLGGRIKDTVLQSLMNEVCITLRTQCGQLYA-----YDWISIPLYTVQVTV 242
Db 606 WRGPACRPP---DSSYN---SLYSTCLELFFKFTIGMGDLFTENYDFKAVFILLIAYVI 659
Qy 243 AVYSFPLACLIGRQFLNPKDYPGHE-----MDLVVPVFTILQFLFWMGLKVA 291
Db 660 LTYILLNMLIALMGETVKNIAQESKNWIKLQRAIFILDTEKSFCLKMRAFRSG--KLL 717
Qy 292 EQLINPFGEDDDF-----ETNWI-----IDRLQVSLLS--VDGMH 326
Db 718 QVGYTPDGKDDVRCFRVDEVNWTNTVNGIINEDPGNCEGVKRTLSLSLRSSRVSGRH 777
Qy 327 QNLPPMERDMYNEAAPQPYTAASARRHSFWGST-----FNISLKEDLEMSKEAD 382
Db 778 -----WKNFALVPLLRASARDQSAQPEEVYLRQFSGSLKPEDAEVFKSPAAS 826
Qy 383 TDK 385
Db 827 GEK 829

APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent in version 3.1
SEQ ID NO 3280
LENGTH: 771
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3280

Query Match 3.2%; Score 94.5; DB 9; Length 771;
Best Local Similarity 20.6%; Pred. No. 5.9; Indels 149; Gaps 25;
Matches 95; Conservative 61; Mismatches 157

QY 139 ILASISVSVKRPPTLHLVLAGFMTHGKHQKQLGLPHNTFWVPWVPMFANLSMKAYLG 198
Db 94 IPRFLDNTSVSV-CLYVMSVWVLLYPEDQOP--LRTAHEIY----VRYKELTKAIVLA 146
QY 199 GRIRDTVLLQSLMNEVC--TLRTQCGOLYAYD--WISIPLVYTVQVTVVYSEFLACL-- 252
Db 147 IRLNDTOLIKSDINATSDPVLKQMAFLVARQOIWLDMP-----EEENTSFMOCLNN 199
QY 253 --IGRQFLPNKDYPGHEMDLVVP-----VFTILOFLFYMGWLKV-----AEQLINP 297
Db 200 ITISKHPKS-----LGKELDILEPKNPEDIYKTHLESSRGAGLTNVDASRHNLASAFVNA 254
QY 298 -----FGEDD-----DDPETNWI-----IDR----- 313
Db 255 FVNAAGFGKDNMLGDDDDKGSWIKTKDGMSTTASFGMLLSKDDTGLDRIDKFTYATE 314
QY 314 -NLQVSLLSVDGMHQNLPMPERD-----MYWNEAAPPPVTAASARSRHSFMGST 363
Db 315 EQIKAGALLGTGTGSGVRLEADPVLALLGDPEVWQERS--VPRVAAIMGLGLAYAGSN 372
QY 364 FNIISKEDLELWSK--BEADTKKESGYSS-TTG-CFLGLQPKNYHLPLKDLTKLLCS 419
Db 373 -----KELLEILLPVVEDVSLDMQLSAMAASVSLGLIFVG---SSNHQVSEAIATILM-- 422
QY 420 KNPILLEGGCKDANQKQKDVH-KFKG-----LDPLKCVPRFKRGSHC 461
Db 423 -----DEERQKHLKDKWTRFMALGLALLYFGROEVDVILDKAVDH----- 465
QY 462 GPQAPSSHTQSAPSSSDTGDPSTDYQIECHMKKKTVEFN 503
Db 466 -----PMKPTSVLASVCWAGTGTVVKLQELLHICNDLIEEN 503

Search completed: July 10, 2003, 12:28:59
Job time : 23.0439 secs

APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent in version 3.1
SEQ ID NO 3280
LENGTH: 771
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3280

Query Match 3.2%; Score 94.5; DB 9; Length 586;
Best Local Similarity 18.8%; Pred. No. 3.9;
Matches 88; Conservative 55; Mismatches 125; Indels 201; Gaps 23;

QY 42 LYSIRGLVRLVMSDDQQLLPEKIALYCDSYIQLIPISVLGYFTLVVSRWMSQYENLP 101
Db 157 LFVFIAGIVFLHSSNKSLYSGRVIFCLDY-----IIFTLRIHIFTVSR-----NL- 204
QY 102 WPDRLMIQVSSFVEGKDEGRLLRLIRYAILGQVLLRLRSISTSVYKRPPTLHLVL-- 159
Db 205 GPKIMLQRTSIEMSSSGSSIPTLRPFEP-----VVLQISIGT-----SSHEVMLS 253
QY 160 -----AGFMTHGKHQKQLGLPHNTFWVPWVPMFANLSMKAYL 197
Db 254 DRCVLLPVLPGVGGVLCVARQILRQNEOR-----WRWIFRSVIEPYL 298
QY 198 G--GRIRDTVLLQSLMNEVC--TLRTQCGOLYAYD--WISIPLVYTVQVTVVYSEFLACL-- 252
Db 299 AMFGQVPSDVGTYDFAHCTFTGNESKPLCVELDEHNLPRPEWITIFLVCIYMLSTNI 358
QY 240 VTVAVYSFFLACLIGRQFLPNKDYPGHEMDLVVPVFTTILQFLFYMGWLKVAEQLINPFG 299
Db 359 LLVNLVAMFGCVAG-----GLVQVL----- 379
QY 300 EDDDDPETNWIIDNLOVSLLSVDGMHQNLPMPERDWMYNEAAPPPVTAASARSRHSF 359
Db 380 -----ENNLKVS-----QKQKQARELTAKPKYTLAAGPFR--- 414
QY 360 MGSTFNISLKKEDLELWSKBEADTKKESGYSTIGCFLGLQP-----KNYHLPLK 410
Db 415 -----WT-----SAVTAC---LQPARCLPGTGRQGHKISLE 442
QY 411 DLKTKLL-----CS---KNPL-LEGQCKDANQKQKDVH-KFKG-----LDPLKCVPRFKRGSHC 461
Db 443 MHKGKIAEFSGQGHQMATCGQGFKNHLRWGGYTVGTVQENNDQVWFQ 491

RESULT 15
US-10-128-714-3280
Sequence 3280, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:04 ; Search time 16.1337 Seconds
(without alignments)
3283.199 Million cell updates/sec

Title: US-09-622-964-29
Perfect score: 2944
Sequence: 1 MTITYTNKANARLGSPSSL.....EHAESYPYRDEAGTKPVLVE 551

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	828.5	28.1	612	2 T32368	hypothetical prote
2	771.5	26.2	557	2 T32367	hypothetical prote
3	756	25.7	499	2 T27630	hypothetical prote
4	755.5	25.7	405	2 T27971	hypothetical prote
5	749.5	25.5	584	2 T19565	hypothetical prote
6	736.5	25.0	632	2 S44917	ZK688.2 protein -
7	722.5	24.5	1355	2 T28715	hypothetical prote
8	665	22.6	413	2 T21644	hypothetical prote
9	663.5	22.5	523	2 T18782	hypothetical prote
10	658.5	22.4	450	2 T18781	hypothetical prote
11	656.5	22.3	405	2 S42371	hypothetical prote
12	655	22.2	513	2 T24210	hypothetical prote
13	646.5	22.0	459	2 S40708	hypothetical prote
14	619.5	21.0	613	2 T16885	hypothetical prote
15	616	20.9	512	2 T19806	hypothetical prote
16	601	20.4	387	2 H89192	protein F32G8.4 [i
17	601	20.4	400	2 T21670	hypothetical prote
18	599.5	20.4	420	2 B88710	hypothetical prote
19	503	17.1	400	2 T20050	protein C43G2.4 [i
20	495	16.8	444	2 T20048	hypothetical prote
21	487.5	16.6	530	2 T28037	hypothetical prote
22	477.5	16.2	806	2 T15468	hypothetical prote
23	466.5	15.8	411	2 T28038	hypothetical prote
24	412	14.0	434	2 T20922	hypothetical prote
25	114.5	3.9	516	2 A80665	probable membrane
26	109.5	3.7	1707	2 T18951	hypothetical prote
27	109.5	3.7	2606	2 T03159	large tegument pro
28	108.5	3.7	315	2 A80677	probable membrane
29	108.5	3.7	560	2 T07964	(S)-N-methylcocciu

30	108	3.7	489	2 S50396	hypothetical prote
31	103.5	3.5	4092	1 S38128	dynein heavy chain
32	102.5	3.5	331	2 AD2063	hypothetical prote
33	102.5	3.5	362	2 T32242	hypothetical prote
34	102.5	3.5	613	2 G82338	conserved hypotet
35	102.5	3.5	1045	2 T18630	hypothetical prote
36	102.5	3.5	1522	2 T00028	brain-specific ang
37	102.5	3.5	2178	2 S29237	calcium channel pr
38	102.5	3.5	2259	2 S29236	calcium channel pr
39	101	3.4	720	2 G64230	stringent response
40	100	3.4	388	2 T16861	hypothetical prote
41	100	3.4	620	2 S55086	probable membrane
42	99.5	3.4	564	2 A48141	ferric reductase (
43	99.5	3.4	570	2 S60426	probable membrane
44	99.5	3.4	685	2 B82297	c-di-GMP phosphodi
45	99.5	3.4	1027	2 B90527	atp-binding protei

ALIGNMENTS

RESULT 1

T32368
hypothetical protein C01B12.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C/Accession: T32368
R/Scheet, P.; Magg, L.
submitted to the EMBL Data Library, September 1997
A/Description: The sequence of C. elegans cosmid C01B12.
A/Reference number: Z21156
A/Accession: T32368
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-612 <SCH>
A/Cross-references: EMBL:AF025458; PIDN:AAB70976.1; GSPDB:GN00020; CESP:C01B12.3
A/Experimental source: strain Bristol N2; clone C01B12
C/Genetics:
A/Gene: CESP:C01B12.3
A/Map position: 2
A/Introns: 25/3; 60/2; 105/2; 138/3; 212/3; 319/3; 369/2; 467/2; 508/3; 573/1
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 28.1%; Score 828.5; DB 2; Length 612;
Best Local Similarity 38.3%; Pred. NO. 1.7e-59;
Matches 160; Conservative 86; Mismatches 145; Indels 27; Gaps 4;

Qy	1	MTITYTNKANARLGSPSSLLCHRGSTYKLYGFLVFIFYLYSIRGLYRMVLSDDQL	60
Db	1	MTVYSLDVASSSPFCYKLLFRWKSITKSWAELVVMCLYAVLSVYIRCLLTMKORA	60
Qy	61	LFEKALYCDYIQLIPISFVLGVVTLVSVRWVQENLWPDRLMTQVSVFVGKDBE	120
Db	61	TFEDLCIFFDYFNIFITFMGVVNAVFRWQIFDNIGWIDTPCLWIYQYKGETER	120
Qy	121	GRLLRRLIRYAILGQVILRSISTSVYKRPPTLHLVLGFMTHGEHQQLKGLPHNT	180
Db	121	AKCVRRNCIRYSILTQAMVYRDVAASVKRPFTFNHLVTAGLTKEKMAEFESIPSHAK	180
Qy	181	FWVPWFANLSMKAYLGGRIIDTVLQSLMNEVCTLRTOCGQLYAYDWISIPLYTVQV	240
Db	181	YWQPMHMLFSGMITLARDEGMISDIIYVDLMKMRQRPVNILSLTFLDFWVPLVYTVQV	240
Qy	241	TVAVYSPFLACLGRQFLNPN----KDPVPGHEMDLVVVFVTLQFLFYGMKLVKAEQLIN	296
Db	241	HLAVRSFLIALFGRYLHPESNRLNDFK-OTIDLYVDIMSLQLFFIFIGMKNVAEVLIN	299
Qy	297	PFGEDDDDFETNWIIDRNQLVSLSDVMGMHQNLPFMRDMYWNAAPOPPYTAASARSRR	356
Db	300	PLGEDDDDFECNWLDRNLQVGLMVDVAVNRYPTLEKQKQWEDAIAPFLYTAESAMRPL	359
Qy	357	HSFNGSTFNLSLKED-----LELWSKEADTD-----KKESSYSS 392	

[illegible]


```

QY      240 VIVAVTSFYLALGQFLNPNPDIQREHNDLTV...:|:~::~||::|:~::~||:~::~||:~::~||
Db      375 VFLAVRVYPAICLVSRQFLISDMK-SKTQMDWFPIMTVLEFIIVIGNMKVAEVLINPLG 433

QY      300 EDDDDPETWTIDRNLOVLSLDGMHQNLPPMERDWMYNNEAAPQPPTYAASARSR--H 357
Db      434 EDDDDPEVNISIDNNISRGMAIVDTTGHYPDLVDVF-----SDPNLPAYSSENSQIPR 488

QY      358 SFMGSTFNISL 368
Db      489 NLTGSAAKVEL 499

RESULT 15
Tl9806
hypothetical protein C37A5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: Tl9806
R;White, S.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19180
A;Accession: Tl9806
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-512 <Wil>
```


A;Cross-references: EMBL:Z92828; PIDN:CAB07337.1; GSPDB:GN00019; CESP:C37A5.1
A;Experimental source: clone C37A5
C;Genetics:
A;Gene: CESP:C37A5.1
A;Map position: 1
A;Introns: 19/3; 60/2; 106/2; 162/1; 265/3; 313/3; 364/2; 388/1; 467/3
C;Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 20.9%; Score 616; DB 2; Length 512;
Best Local Similarity 28.3%; Pred. No. 2.9e-42;
Matches 161; Conservative 103; Mismatches 200; Indels 104; Gaps 18;
Qy 1 MTITYTNKAVANARIGSFSSLLLCWRGSIYKLLYGELVFIFLYYSIRGLYRMVLSDDQL 60
Db 1 MTVSYNLDVSSVSFPFKLLFRKGSVWKSISLWLLGLYYLVVYRVALTTEQKA 60
Qy 61 LFEKALYCDSYI-QLIPISVLGYVTLVSRWWSQYENLPWDRMLQIVSSPVEGKDE 119
Db 61 GVRKYIDHLDQNLKCVPLTFMLAFVTVIIVDRWKNMFANIGFIENATATATLVKGTG 120
Qy 120 EGRLLRRLIYAILGOVLILRSISTSVYKFPFTLHLVLVLAGFMTHGEHKOLQKLGPHN 179
Db 121 DVLAKRTIIRYLVLTQVLVFRDISLVKRRFPNHDAILKAGFLQ--DHESIILGGDNGR 178
Qy 180 T-FWVPVWVFANLSMKAYLGGRIIRDTVLLQSLMNEVCTLRTOCGOLYAYOWISPLVYTQ 238
Db 179 TNYWPNVWSSAILQKLPEDGNI PAAPLFNSVWQEVKTFRSNMATLCNYDWVPIPIAYPQ 238
Qy 239 VVTVAVSFFLACLIGROFLNPNKDYPGHEMDLVVPVFTILOFLYMGWLKVAQLINPF 298
Db 239 VVFAVRVYFTCLFTROHLMEDT---KTIDYFPILTVPFTFFGMGMKVAEALLNPL 295
Qy 299 GEDDDDFETNWIIDRNQVLSLVSDGMHONLPPMERDMYNEAAPQPPYTAASARS-RRH 357
Db 296 GEDDDDFECNYLIDRNATGMAIVNSKYSDVPEMLADKP-NDPSYAPYPPEKVIDSGADH 354
Qy 358 SFMGSTFNISLKEDLELWSKEADTDKESGYSTICGFLGLQPKNYHLPLKDLTKLL 417
Db 355 ALVGSAGQVTL-----AEPNDI---IDMMKVDL- 379
Qy 418 CSKNPLLEGCKDANKQKQKQVWPKGLDFLKCVPRPKRG---SHCGPQAPS-SHPTEQ 473
Db 380 --NSPIVVGR---RNTNNTSTIRR-----RLSAGRRSHSVQHLGPEKPEPTSPFSQ 427
Qy 474 S-AP-----SSDGTGDPSTDYQEIChMKKKTVE 501
Db 428 SNAPORPYGAFELNGFNLSGLSGITSSQSHLPKLESEETTD-PLPAQOPVFLPRTLSE 486
Qy 502 FNLNIPESPT-----EHLQORRLDQM 522
Db 487 ----EPTSPTLPFDQTLFHVNDNGLNKM 510

Search completed: July 10, 2003, 12:27:12
Job time : 18.467 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 9.11903 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-29

Perfect score: 2944

Sequence: 1 MTITYTNKVNARLGSPSSL.....EHAESYPYRDEAGTKPVLVE 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1908	64.8	585	1 VMD2 HUMAN	O76090 homo sapien
2	756	25.7	499	1 YHDI CAEEL	Q23369 caenorhabdi
3	755.5	25.7	405	1 YS63 CAEEL	Q09379 caenorhabdi
4	736.5	25.0	632	1 Y022 CAEEL	P34672 caenorhabdi
5	665	22.6	413	1 YV4Q CAEEL	Q45435 caenorhabdi
6	663.5	22.5	523	1 Y0E4 CAEEL	Q17529 caenorhabdi
7	658.5	22.4	450	1 Y0E3 CAEEL	Q17528 caenorhabdi
8	656.5	22.3	456	1 YNK4 CAEEL	P34577 caenorhabdi
9	655	22.2	513	1 YXAK CAEEL	Q21973 caenorhabdi
10	646.5	22.0	459	1 YKT8 CAEEL	P34319 caenorhabdi
11	619.5	21.0	501	1 YSV1 CAEEL	Q22566 caenorhabdi
12	601	20.4	400	1 YV6L CAEEL	Q19978 caenorhabdi
13	599.5	20.4	420	1 YCBL CAEEL	Q94175 caenorhabdi
14	546	18.5	884	1 YAVK CAEEL	Q17851 caenorhabdi
15	502.5	17.1	602	1 YHS4 CAEEL	O19303 caenorhabdi
16	466.5	15.8	411	1 YHS5 CAEEL	O18304 caenorhabdi
17	436	14.8	81	1 VMD2 MOUSE	O88870 mus musculu
18	412	14.0	434	1 YSWJ CAEEL	O45363 caenorhabdi
19	108.5	3.7	315	1 YNEE SALTY	Q82706 salmonella
20	108.5	3.7	315	1 YNEE SALTY	Q82706 salmonella
21	108	3.7	489	1 YN20 YEAST	P40210 saccharomyc
22	104.5	3.5	389	1 O95C DROME	Q9VH62 drosophila
23	103.5	3.5	4092	1 DYHC YEAST	P36022 saccharomyc
24	102.5	3.5	613	1 UBID VIBCH	Q9KV48 vibrio chol
25	102.5	3.5	1522	1 BA13 HUMAN	O60242 homo sapien
26	102.5	3.5	2259	1 CCAF RABIT	Q02343 oryctolagus
27	101.5	3.4	306	1 Y114 FALSO	Q8XY11 raistonia s
28	101	3.4	720	1 SPOT MYCCE	P47520 mycoplasma
29	100	3.4	620	1 YN20 YEAST	Q03162 saccharomyc
30	99.5	3.4	564	1 FRP1 SCHPO	Q04800 schizosacch
31	99.5	3.4	570	1 YGR0 YEAST	P53109 saccharomyc
32	98.5	3.3	802	1 YGN9 YEAST	P53121 saccharomyc
33	98	3.3	4447	1 PKSK_BACSU	P40803 bacillus su

34 95.5 3.2 347 1 NU2M HIPAM Q9ZZZ0 hippopotamu
35 94 3.2 487 1 USID_HAEDU Q9RPX1 haemophilus
36 93.5 3.2 299 1 HSF6_ARATH Q98CW4 arabidopsis
37 93 3.2 374 1 Y006 BORBU O51039 borrelia bu
38 93 3.2 459 1 NU4M_PELSU O79677 pelomedusa
39 92.5 3.1 199 1 YRF2_SHIFL P37788 shigella fl
40 92.5 3.1 679 1 YKR9_YEAST P34237 saccharomyc
41 92.5 3.1 1663 1 HAPD_HUMAN P34237 saccharomyc
42 92.5 3.1 1862 1 ANK1_MOUSE O60229 homo sapien
43 92 3.1 492 1 Y821_RICPR Q02357 mus musculu
44 91.5 3.1 669 1 YH06_HAEN P45335 haemophilus
45 91 3.1 347 1 NU2M_PIG O79875 sus scrofa

ALIGNMENTS

RESULT 1
VMD2 HUMAN
ID VMD2 HUMAN STANDARD; PRT; 585 AA.
AC O76090; O75904;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bestrophin (Viteliform macular dystrophy protein). (TUI58).
GN VMD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS BMD.
RX MEDLINE=98367043; PubMed=3700209;
RA Marquardt A., Stoehr H., Passmore L.A., Kraemer F., Rivera A.,
RA Weber B.H.F.;
RT "Mutations in a novel gene, VMD2, encoding a protein of unknown
RT properties cause juvenile-onset vitelliform macular dystrophy (Best's
RT disease).";
RL Hum. Mol. Genet. 7:1517-1525 (1998).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND E-299.
RX MEDLINE=98324772; PubMed=9662395;
RA Petrukhin K., Koisti M.J., Bakall B., Li W., Xie G., Marknell T.,
RA Sandgren O., Forsman K., Holmgren G., Andreasson S., Vujic M.,
RA Bergen A.A.B., McGarty-Dugan V., Figueroa D., Austin C.P.,
RA Metzker M.L., Caskey C.T., Wadelius C.;
RT "Identification of the gene responsible for Best macular dystrophy.";
RL Nat. Genet. 19:241-247 (1998).
RN [3]
RP VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.
RX MEDLINE=99265978; PubMed=10331951;
RA Caldwell G.M., Kakuk L.E., Griesinger I.B., Simpson S.A., Nowak N.J.,
RA Small K.W., Maumenee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,
RA Ayagari R.;
RT "Bestrophin gene mutations in patients with Best vitelliform macular
RT dystrophy.";
RL Genomics 58:98-101 (1999).
RN [4]
RP VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.
RX MEDLINE=99320852; PubMed=10394929;
RA Bakall B., Marknell T., Ingvaest S., Koisti M.J., Sandgren O., Li W.,
RA Bergen A.A.B., Andreasson S., Rosenberg T., Petrukhin K., Wadelius C.;
RT "The mutation spectrum of the bestrophin protein -- functional
RT implications.";
RL Hum. Genet. 104:383-389 (1999).
RN [5]
RP VARIANTS AND/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.
RX MEDLINE=99381534; PubMed=10453731;
RA Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,
RA Sharma S., Gerrard B., Li W., Metzker M.L., Wadelius C., Caskey C.T.,
RA Dean M., Petrukhin K.;
RT "Evaluation of the Best disease gene in patients with age-related
RT macular degeneration and other maculopathies.";

Hum. Genet. 104:449-453 (1999).

[6] VARIANTS BMD P-16; C-17; N-73; H-92; C-218; H-218; L-235 AND S-296.
MEDLINE=21138457; PubMed=11241846;
Marchant D., Gogat K., Boutboul S., Pequignot M., Sternberg C.,
Dureau P., Roche O., Uteza Y., Hache J.C., Puech B., Puech V.,
Dumur V., Mouillon M., Munier P.L., Schorderet D.F., Marsac C.,
Dufier J.L., Abitbol M.;
identification of novel VMD2 gene mutations in patients with Best
vitelliform macular dystrophy.";
Hum. Mutat. 17:235-235 (2001).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN VMD2 ARE THE CAUSE OF BEST MACULAR DYSTROPHY
(BMD); ALSO KNOWN AS VITELLIFORM MACULAR DYSTROPHY TYPE 2. BMD IS
AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY TYPICAL "EGG-YOLK"
MACULAR LESIONS DUE TO ABNORMAL ACCUMULATION OF LIPOFUSCIN WITHIN
AND BENEATH THE RETINAL PIGMENT EPITHELIUM CELLS. PROGRESSION OF
THE DISEASE LEADS TO DESTRUCTION OF THE RETINAL PIGMENT EPITHELIUM
AND VISION LOSS.
CC -!- DISEASE: DEFECTS IN VMD2 COULD BE THE CAUSE OF OTHER FORMS OF
MACULOPATHY AS BULL'S EYE MACULOPATHY, AND ADULT VITELLIFORM
MACULAR DEGENERATION (AMD).
CC -!- DISEASE: IN RARE CASES, DEFECTS IN VMD2 MAY INCREASE
SUSCEPTIBILITY TO AGE-RELATED MACULAR DEGENERATION (AMD).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -!- DATABASE: NAME=VMD2 mutation database;
WWW="http://www.uni-wuerzburg.de/humangenetics/vmd2.html".
CC -!- DATABASE: NAME=Mutations of the VMD2 gene;
NOTE=Retina International's Scientific Newsletter;
WWW="http://www.retina-international.com/sci-news/vmd2mut.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AF073500; AAC64926.1; JOINED.
DR EMBL; AF073491; AAC64926.1; JOINED.
DR EMBL; AF073492; AAC64926.1; JOINED.
DR EMBL; AF073493; AAC64926.1; JOINED.
DR EMBL; AF073494; AAC64926.1; JOINED.
DR EMBL; AF073495; AAC64926.1; JOINED.
DR EMBL; AF073496; AAC64926.1; JOINED.
DR EMBL; AF073497; AAC64926.1; JOINED.
DR EMBL; AF073498; AAC64926.1; JOINED.
DR EMBL; AF073499; AAC64926.1; JOINED.
DR EMBL; AF057169; AAC64343.1; --
DR EMBL; AF057170; AAC64344.1; --
DR EMBL; AF073501; AAC33766.1; --
DR Genew; HGNC:12703; VMD2.
DR MIM; 153700; --
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01462; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Alternative splicing; Disease mutation: Polymorphism; Vision.
VARSPPLIC 291 483
ASQINPFGEDEDDDETTNVIYDNLQVSLAVDEMHDPLRP
MEFDPMYKPEPPPTTAASQAQFRASPGWSTNLSNKE
MEFQPNQDEEDDAHAGIIGRFLGSHDHPFRANSRYKLL
WPKRESLLHGLPKKHKAQKQNVGQEDNKAWKLAVDAFK
SAPLYPGVYSAPOTPLSPFPMFPLEP -> GLSRALLG
WPHGQRGQQLLETRMOCQKRVSRVSSQAQWRTVPVPA
TREAEAGSLEPGRRLWQSSSTPLERMLRLPTGLST
GICRCPWLWNRCTRCLGWSRTCLGSPSHSPPTQLLPPS
SVEPLWAPPST (IN ISOFORM 2).
T -> P (IN BMD AND AVMD).
/FTid=VAR_000830.
V -> A (IN BMD).
/FTid=VAR_000831.
V -> M (IN BMD)

Hum. Genet. 104:449-453 (1999).

[6] VARIANTS BMD P-16; C-17; N-73; H-92; C-218; H-218; L-235 AND S-296.
MEDLINE=21138457; PubMed=11241846;
Marchant D., Gogat K., Boutboul S., Pequignot M., Sternberg C.,
Dureau P., Roche O., Uteza Y., Hache J.C., Puech B., Puech V.,
Dumur V., Mouillon M., Munier P.L., Schorderet D.F., Marsac C.,
Dufier J.L., Abitbol M.;
identification of novel VMD2 gene mutations in patients with Best
vitelliform macular dystrophy.";
Hum. Mutat. 17:235-235 (2001).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN VMD2 ARE THE CAUSE OF BEST MACULAR DYSTROPHY
(BMD); ALSO KNOWN AS VITELLIFORM MACULAR DYSTROPHY TYPE 2. BMD IS
AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY TYPICAL "EGG-YOLK"
MACULAR LESIONS DUE TO ABNORMAL ACCUMULATION OF LIPOFUSCIN WITHIN
AND BENEATH THE RETINAL PIGMENT EPITHELIUM CELLS. PROGRESSION OF
THE DISEASE LEADS TO DESTRUCTION OF THE RETINAL PIGMENT EPITHELIUM
AND VISION LOSS.
CC -!- DISEASE: DEFECTS IN VMD2 COULD BE THE CAUSE OF OTHER FORMS OF
MACULOPATHY AS BULL'S EYE MACULOPATHY, AND ADULT VITELLIFORM
MACULAR DEGENERATION (AMD).
CC -!- DISEASE: IN RARE CASES, DEFECTS IN VMD2 MAY INCREASE
SUSCEPTIBILITY TO AGE-RELATED MACULAR DEGENERATION (AMD).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -!- DATABASE: NAME=VMD2 mutation database;
WWW="http://www.uni-wuerzburg.de/humangenetics/vmd2.html".
CC -!- DATABASE: NAME=Mutations of the VMD2 gene;
NOTE=Retina International's Scientific Newsletter;
WWW="http://www.retina-international.com/sci-news/vmd2mut.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AF073500; AAC64926.1; JOINED.
DR EMBL; AF073491; AAC64926.1; JOINED.
DR EMBL; AF073492; AAC64926.1; JOINED.
DR EMBL; AF073493; AAC64926.1; JOINED.
DR EMBL; AF073494; AAC64926.1; JOINED.
DR EMBL; AF073495; AAC64926.1; JOINED.
DR EMBL; AF073496; AAC64926.1; JOINED.
DR EMBL; AF073497; AAC64926.1; JOINED.
DR EMBL; AF073498; AAC64926.1; JOINED.
DR EMBL; AF073499; AAC64926.1; JOINED.
DR EMBL; AF057169; AAC64343.1; --
DR EMBL; AF057170; AAC64344.1; --
DR EMBL; AF073501; AAC33766.1; --
DR Genew; HGNC:12703; VMD2.
DR MIM; 153700; --
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01462; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Alternative splicing; Disease mutation: Polymorphism; Vision.
VARSPPLIC 291 483
ASQINPFGEDEDDDETTNVIYDNLQVSLAVDEMHDPLRP
MEFDPMYKPEPPPTTAASQAQFRASPGWSTNLSNKE
MEFQPNQDEEDDAHAGIIGRFLGSHDHPFRANSRYKLL
WPKRESLLHGLPKKHKAQKQNVGQEDNKAWKLAVDAFK
SAPLYPGVYSAPOTPLSPFPMFPLEP -> GLSRALLG
WPHGQRGQQLLETRMOCQKRVSRVSSQAQWRTVPVPA
TREAEAGSLEPGRRLWQSSSTPLERMLRLPTGLST
GICRCPWLWNRCTRCLGWSRTCLGSPSHSPPTQLLPPS
SVEPLWAPPST (IN ISOFORM 2).
T -> P (IN BMD AND AVMD).
/FTid=VAR_000830.
V -> A (IN BMD).
/FTid=VAR_000831.
V -> M (IN BMD)

```

Query Match      64.8%; Score 1908; DB 1; Length 585;
Best Local Similarity 64.5%; Pred. No. 4e-147;
Matches 379; Conservative 59; Mismatches 100; Indels 50; Gaps 10;

QY 1 MTITTYKVNARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60
DB 1 MTITTSQVANARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60

QY 61 LPEKALYCDSYIQLIPISFVLGFYVTVLVSRWMSQYENLPWPDRLMIQVSSFVEGKDE 120
DB 61 MEKALTLYCDSYIQLIPISFVLGFYVTVLVSRWMSQYENLPWPDRLMSLVSGFVEGKDEQ 120

QY 121 GRLLRRTLIYAILGOVLIIRLSISTSVYKRPFTLHLVLVLAGFMTHGEHKQLKGLPHNT 180
DB 121 GRLLRRTLIYANLGNVLIRLSVAVYKRPFSQAHLVQAGFMTPAETHKQLEKLSLPHNM 180

QY 181 FWPVWVWFANLWKAVLGRIRDTVLQSLMNEVCTLRTOCQOLVAYDWISIPLVYTVQV 240
DB 181 FWPVWVWFANLWKAVLGRIRDTVLQSLMNEVCTLRTOCQOLVAYDWISIPLVYTVQV 240

QY 241 TVAVYSFFLACILGRQFLNPKNDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINPGE 300
DB 241 TVAVYSFFLACILGRQFLNPKNDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINPGE 300

QY 301 DDDDPETNWIIDRNQVLSLVSDGKHONLPPMERDMYNEAAPQPPYTAASARSRRHSPM 360
DB 301 DDDDPETNWIIDRNQVLSLVSDGKHONLPPMERDMYNEAAPQPPYTAASARSRRHSPM 360

QY 361 GSTFNISLKDELELWSKEADTDKESGYSTIGCFGLQPKVHPLKDKLTKLCSK 420
DB 361 GSTFNISLKDELELWSKEADTDKESGYSTIGCFGLQPKVHPLKDKLTKLCSK 420

QY 421 NPLL--EGCKD-----ANQNKQD--VWKFGLDPLKCVPRFGRSGHCGQAPSS--- 468
DB 417 RESLLHEGLPKNHKAQNVQGEONKAWKLAVDAFKASAPLYQPGYYSAQTPLSPTP 476

QY 469 --HPTQSPSS--SDTG-----DQSTDYQIICHMKKT 499
DB 477 MEFFLEPSAPSLHSVTGTDKSLKTVSSGAKSFELLSGSDGALMEHPEVSQVRRT 536

QY 500 VEFNL-NIPESPTLQORLDQMSNIOALMKEHAEYS---PYDEA 543
DB 537 VEFNLTDMEIPEINHLKE-PLSQSPNIIHTLKDHWDPTWALENDEA 583

```

```

RESULT 2
YHDI_CABEL STANDARD; PRT; 499 AA.
AC Q23369;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZC518.1 in chromosome IV.
GN ZC518.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Thomas K.;
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z68753; CAA92989.1; -.
DR WormPep; ZC518.1; CE06601.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD02802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 499 AA; 57229 MW; 4E87725437A5C9DF CRC64;

Query Match      25.7%; Score 756; DB 1; Length 499;
Best Local Similarity 34.9%; Pred. No. 9.7e-54;
Matches 176; Conservative 87; Mismatches 183; Indels 58; Gaps 10;

QY 1 MTITTYKVNARLGSFSLLCWRSIYKLYLGRFLVFIFLYYSIRGLYRMVLSDDOOL 60
DB 1 MTISYTLDSVQTNLQSPFSLLRWRGYSVWKAQVQLAVTAVFLLISCIYMLSPSQD 60

QY 61 LPEKALYCDSYIQL-IPISFVLGFYVTVLVSRWMSQYENLPWPDRLMIQVSSFVEGKDE 119
DB 61 VPEQLIRYFNKLDANIPLTLLGFFSVFVVARWGSILNGIWDASLLFATYIRGADE 120

QY 120 EGRLLRRTLIYAILGOVLIIRLSISTSVYKRPFTLHLVLVLAGFMTHGEHKQLKGLPHN 179
DB 121 ETRVIRRNLYVLVLSQALVLRDISMQVRKFTMDTLAASGLMTHSEMDILDKDPYS 180

QY 180 TFWPWWVWFANLWKAVLGRIRDTVLQSLMNEVCTLRTOCQOLVAYDWISIPLVYTVQV 239
DB 181 RYMTSIQWLSNLVTECKGKGVSYLLMNKIVDEIGKFRHGLASLLKYDWDVPLVYPQV 240

QY 240 TVAVYSFFLACILGRQFL--NPNKDYPGHEMDLVVPVFTILQFLFYMGWLKVAEQLINP 297
DB 241 IFLAVRIYFMICLGRQFIVTGNPS-----GIDLWLPITTVQVFLVYMGWVKAELNLP 296

QY 298 FGEDDDPETNWIIDRNQVLSLVSDGKHON-----LPPMERDMYNEAAP 343
DB 297 LGEDDDDLCEYIIDKNLITGLSIVDTWVKHDDTGYSMVEBMAKTPAQKDEFGWIDKI 356

QY 344 QPPYTAASARSRRHSPFSGSTFNISLKDELELWSKEADTDKESGYSTIGCFGLQPK 403
DB 357 APLYSMESAERSVHPLVGSASKINLVK-----NKKE-----IVMTTPH 393

QY 404 NVHLPLKDLKTKLCSKNLLEGCKDQANQKNQKDVWKFGLDPL-----KCVPRFGRSGS 459
DB 394 KNLSELDPSEQ----KTYLRRVNVSDHNAKHAQK----RGLERANSPDKCLSKVRSRN 445

QY 460 HCGPQAPSSHPTQSPSSSDTG 483
DB 446 --GKFRTSANGSQNGVDLWTRAGD 467

RESULT 3
YS63_CABEL STANDARD; PRT; 405 AA.
ID YS63_CABEL
AC Q09379;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK675.3 in chromosome II.
GN ZK675.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Sims M.;
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Z46812; CAA86845.1; -
 DR WormPep; ZK675.3; CE01722.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 405 AA; 47835 MW; 0F589D874B9E0E61 CRC64;

Query Match 25.7%; Score 755.5; DB 1; Length 405;
 Best Local Similarity 41.7%; Pred. No. 8.1e-54;
 Matches 150; Conservative 77; Mismatches 118; Indels 15; Gaps 4;

QY 1 MTITYNKVANARLGSFSSLLCWRGSIYKLLYGEFLVFIFYYSIRGLYRMVLSDDQL 60
 DB 1 MTISYSD-----TFKLFRWKGLKAIWKHLIFLTMYIINAYYRFGWTKQQN 52
 QY 61 LPEKALYCDYSYQIOLIPISFVLGFWYLVVSRWWSQYENLPWDRMLQVSSFVEGKDEE 120
 DB 53 EPIKYVNLVDGWTKIEIFLTFGLFYVAMIVRWWDCCOLISWPDHLLYNVSALIRGDPE 112
 QY 121 GRLLRRLIRVAILGQVLILRSISTSVYKRPPTLHLVLVLAGFMTHGHEKQKLG---P 177
 DB 113 TRIIRKIARTILTSVLWRSISLRVLARYPTDDHLVDSGLMTKEENWFKSLVHVDP 172
 QY 178 HNTFWPVMVPMANLSKRAYLGRIRDTVLLQSLMNEVCTLRTOGQLYAYDWISIPLYVT 237
 DB 173 HQKWWPLNWTQTMWRCFEKGTLLTHNELRVLLDALEKYRNGFPQLFYDWIAIPLYVT 232
 QY 238 QVTVVAVYSFLACLGROFLNPKDYPGHEMDLVVPFTILOFLYGMWLKVAEQLNP 297
 DB 233 QVSTISVYGYFLPALIGROY--PSKNEEIEIVDVYVPIFTILOFLYGMWLKVAEQLNP 290
 QY 298 FGEDDDFETNWIIDRNLOVSLLSVDGMHQLNPPMERDMYNEAAPPYTAASARRRH 357
 DB 291 FGADDEDFEYNIILERNLEVSLVDELHNQVPPYVESLDEI--RLHTSASSKLSNH 348

RESULT 4
 Y022 CAEEL STANDARD; PRT; 632 AA.
 AC P34672;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK688.2 in chromosome III.
 GN ZK688.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Fraser A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kiraten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).

CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; L16621; AAA28228.1; -
 DR PIR; S44917; S44917.7
 DR WormPep; ZK688.2; CE00460.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 73829 MW; 0ABDD1755EF11642 CRC64;

Query Match 25.0%; Score 736.5; DB 1; Length 632;
 Best Local Similarity 38.6%; Pred. No. 5.1e-52;
 Matches 142; Conservative 82; Mismatches 143; Indels 1; Gaps 1;

QY 1 MTITYNKVANARLGSFSSLLCWRGSIYKLLYGEFLVFIFYYSIRGLYRMVLSDDQL 60
 DB 1 MTINYNLAVSTSKPWTFLFKLLKWRGSIWKAVILELAVLWLVYLSVIRTALNPGQOR 60
 QY 61 LPEKALYCDYSYQIOLIPISFVLGFWYLVVSRWWSQYENLPWDRMLQVSSFVEGKDEE 120
 DB 61 TFERIVQYCDUSRLSYPLNPLMGFFVTAVVNRVYLYQIIFIDNIGLMAEYVGRTEQ 120
 QY 121 GRLLRRLIRVAILGQVLILRSISTSVYKRPPTLHLVLVLAGFMTHGHEKQKLGPHNT 180
 DB 121 ARVYRNIVRYCYLAQVLVFRDISMRTRRPFTLDTVVAAGFWMPHEKDFEIQYKYSK 180
 QY 181 FWTWVWFANLSKAYLGRIRDTVLLQSLMNEVCTLRTOGQLYAYDWISIPLYVTQV 240
 DB 181 YWYFPQWAFSLTYEARCKGLIESDYQYVVQVDEIKKFTGLAWICNDYDWPPIPIYQV 240
 QY 241 TVAVYSFLACLGROFLNPKDYPGHEMDLVVPFTILOFLYGMWLKVAEQLNPFG 300
 DB 241 CLAVHTYFLVCLLARQYVVSSEHADNKTIEDLYFPINSTLOFIFYGMWLKVAEQLNPFG 300
 QY 301 DDDDFETNWIIDRNLOVSLLSVDGMHQLNPPMERDMYNEAAPPYTAASARRRHIFM 360
 DB 301 DDDDFECNALIDRNITVLMVMDGYDRAPDLKRDDEFWDEBEV-EPLYSEETAKIPNPLK 359
 QY 361 GSTFNISL 368
 DB 360 GSVSDVKL 367

RESULT 5
 YV4Q CAEEL STANDARD; PRT; 413 AA.
 AC O45435;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein F32B6.9 in chromosome IV.
 GN F32B6.9
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Basham V.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

DR WormPep; B0564.3.; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 450 AA; 53275 MW; A0FED9AA76166AD7 CRC64;

Query Match      22.4%; Score 658.5; DB 1; Length 450;
Best Local Similarity 34.3%; Pred. No. 6.9e-46;
Matches 141; Conservative 76; Mismatches 165; Indels 29; Gaps 5;

QY 1 MTITNTKVNARLGSFSSLLLCWRGSIYKLLYGEFLVFIFLYSIRGLYRWLVSSDQQL 60
DB 1 MTINYHKEITSHPTWTFLLFKWGSIKAVYMETIIFLCYIGIISVIKTAGESSOR 60
QY 61 LFEKALYCDYSIOLIPISFVLGFFVTLVSRWMSQYENLPWDRMLQVSSVEGKDEE 120
DB 61 TFEELVRYFDKRLSYIPLFVLGFFVTVVNRWTKLYQTIGFDNVGLMANCYIRGATEK 120
QY 121 GRLLRRLIRYAILGQVLIILRSISTSVYKRPPTLHLVLAGEFTHGE-----HKQLQ 172
DB 121 ARIYRRNIMRYCELVQILVFRDMSMTRRRPPTMETVVAAGFMKHELYNSYDTKYSN 180
QY 173 KGLPHNTFWPVPWFANLSMKAYLGGRIKRTDVLLOSIMNEVCTLTCCGOLYAYDWISI 232
DB 181 KLG---TKYWIIPANWALCMYTKARKDGYIBSDYFKAQMEGEIRTWRTNIWVCYNDWVPL 237
QY 233 PLVYTQVTVVAVYFFFLACLGROFLNPKDYPGHEMDLVVVFVFTILOFLFYMGWLKVAE 292
DB 238 PLMPYQLVCLAVNLVFLVSIARQ-LVIEKHGMVDEVDVYFPVMTFLOFIYNGWLKVID 296
QY 293 QLINPFGEDDDFETNWIIDRLNQLVSLSDVGMHQNLPMPERDMYVNEAAPQPPYTAASA 352
DB 297 VMLNPFGEEDDDFETNALIDRNITMGLMIADN-PNSTELKDKDFYDEVDPVLYSEBS 355
QY 353 RSRHSPMGSTFNISLKED-----LELWSKEEADYTKKE 387
DB 356 NIPNHHYGSVSEVLEQKGNAPVMMPHSQSAANLRMRMGSFKSVDEDEKD 406

RESULT 8
YXAK CAEEL STANDARD; PRT; 456 AA.
ID YXAK CAEEL
AC P34577.
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T20G5.4 in chromosome III.
GN T20G5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berka M., Smith A., Lloyd C.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z30423; CAA83005.2; -.
DR WormPep; B0564.3.; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 456 AA; 53136 MW; EDE22B541307F3C8 CRC64;

Query Match      22.3%; Score 656.5; DB 1; Length 456;
Best Local Similarity 35.2%; Pred. No. 1e-45;
Matches 130; Conservative 85; Mismatches 151; Indels 3; Gaps 3;

QY 1 MTITNTKVNARLGSFSSLLLCWRGSIYKLLYGEFLVFIFLYSIRGLYRWLVSSDQQL 60
DB 1 MTVSYNQSVATSRPMTFLALIFRWGRSVMWSAIQYSVMGLYFLVSAIRFLSAYQQQ 60
QY 61 LFEKALYCDYSIOLIPISFVLGFFVTLVSRWMSQYENLPWDRMLQVSSVEGKDEE 120
DB 61 IFVRLVDYVNSRMSYVPLDWMGLGFFIAGVLRFRWLYDIIGFDNIACSTATVIRGDSR 120
QY 121 GRLLRRLIRYAILGQVLIILRSISTSVYKRPPTLHLVLAGEFTHGEHKLQKLGPHNT 180
DB 121 AKQYRNNIIRYCELTQVLIIFRDLMSKARKRFPFLDTVAAGFMMPHEKANFDLIQYNNK 180
QY 181 FWPVWTFANLSMKAYLGGRIKRTDVLLOSIMNEVCTLTCCGOLYAYDWISIPLVTVQV 240
DB 181 YFLPFWNALVYNARKEGLIEGDIYVTVISEDIKKFRTGLAWVCYNDWVPLIIPVTV 240
QY 241 TVAVYFFFLACLGROFLNPKDYPGHEMDLVVVFVFTILOFLFYMGWLKVAEOLINPFE 300
DB 241 CLAVHMYFFVGLIARQYVKGSEIDP-DMIDLVPFMTSICQFVFMGLVKGEGLLNPWE 299
QY 301 DDDDFETNWIIDRLNQLVSLSDVGMHQNLPMPERDMYVNEAAPQPPYTAASARR-HSF 359
DB 300 DPDDFETNMLIDRLNGLKIVDEGYDKTPLEKDAFMDDTW-VPLYSEASAEKRYHQ 358
QY 360 MGSTFNISL 368
DB 359 QGSLAHIKI 367

RESULT 9
YXAK CAEEL STANDARD; PRT; 513 AA.
ID YXAK CAEEL
AC Q21973.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein R13.3 in chromosome IV.
GN R13.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Matthews P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73105; CAA97442.1; -.
DR WormPep; R13.3; CR06320.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.

```


[illegible]

```

[2]
RN REVISIONS.
RP Durbin R.;
RA Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL CLC SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; 272509; CAA96648.2; --
CC WormPep; F32G8.4; CE23700.
CC InterPro; IPR000615; Worm_fam_8.
CC Pfam; PF01062; DUF289; 1.
CC ProDom; PD002802; Worm_fam_8; 1.
CC Hypothetical protein.
CC KW SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B862E84 CRC64;
-----
Query Match 20.4%; Score 601; DB 1; Length 400;
Best Local Similarity 34.9%; Pred. No. 2.7e-41;
Matches 134; Conservative 66; Mismatches 150; Indels 34; Gaps 9;
-----
Qy 1 MTITYNKVANRLGFSFSSLLLCWRGSIYKLLYGEFLVFIFYLSIRGLYRMVLSDDQL 60
Db 1 MTISYDDE-----FSSLMRLWRGSIWKAVLKDLGFIAYIIVLAFQWYLLDEKGKE 52
Qy 61 LFEKALYCDSYQLIPISVLGPFYTVLVSRRWMSQVENLPWDRLMIOVSSFVEGKDEE 120
Db 53 YFTGWINWCISGIAQYPLSLFELGFFVSLIVARWWEQFNCSWPDGXMMIVSACLPG--NE 110
Qy 121 GLRLRTLRIRYAILGQVLIILRSISTSVYKFPPTLHLVLVLAGFMTHGEHKQLKGLPHNT 180
Db 111 NMVVRQTIARWSSLOAAIAWSGVSVKTLKFFTERHNVASKLMTTEEYDLYMTNDAPHGK 170
Qy 181 FWPVWVFWANLSMKAYLGIRDTVLVLSQSLMNEVCTLRTCGQOLYADWTISPLVYTVQV 240
Db 171 WFIPLMTVNLKIKQKQKQ--IIDSQMDMLKQVYSVRDGFAMLVFVDWIKIPLVYTVQV 229
Qy 241 TVAVYSFFLACLIGRQFLNPNKOYFGHEMDLV--PVFTILOFLFYMGWLKVASQLINPP 298
Db 230 AIATYGVFFCLIGRQ--PKLDQRSMEKETILFPITFTTQMLFYLGLWLVKGQ-----F 281
Qy 299 GEDDDDDFTNWIIDRNLOVSLLSVDGHQNIPLPNERDMYWNAAAPQPPYTAAS----- 351
Db 282 SIRKIKISELNYVLDRNTAIAHMAASELSDQLPSGAPM-----VPAVPHTRASFKIQDVI 336
Qy 352 ARSRHSPMGSTFNISL--KKEDLE 374
Db 337 PKSLHAGFKLSEAEWKLKPEDLE 360
-----
RESULT 13
YCEL CAEEL
ID YCEL_CAEEL STANDARD; PRT; 420 AA.
AC Q94175;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C43G2.4 in chromosome IV.
GN C43G2.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton B.; Wohlmann P.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

```

```

CC CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; U70848; AAB09111.1; -.
CC DR WormPep; C4362.4; CE08693.
CC DR InterPro; IPR000615; Worm_fam_8.
CC DR Pfam; PF01062; DUF289; 1.
CC DR ProDom; PD002802; Worm_fam_8; 1.
CC KW Hypothetical protein.
CC SQ SEQUENCE 420 AA; 50223 MW; 5078D15E1D414A26 CRC64;
CC
Query Match 20.4%; Score 599.5; DB 1; Length 420;
Best Local Similarity 36.0%; Pred. No. 3.9e-41;
Matches 133; Conservative 67; Mismatches 124; Indels 45; Gaps 11;
QY 1 MTITYNKVANRLGSSLLLCWRGSIYKLYGELFVIFLYSIR-----GLYRWVLSS 56
DB 1 MTISYS-----GNFFRLLRKWSIWRSVWRELFLFLFVPIRFSAPHFFNYDPT 52
QY 57 DQ---QLLFEKALYCDSDYIOLIPISFVLGFPVTLVSRWMSQYENLPWDLMLQVSS 113
DB 53 DSKGVRKLPKVCNEFHEYTKMIPITFLGFGVSVNVRWRQFETLRWPEDFLSILCLL 112
QY 114 VEGDEEGRLLRRTLRVAILGQVLIRSIYSTVYKRPFTLHLVLAFGFMTHGEHKQLQK 173
DB 113 LPSK--ESRPARHQIARYNLTSALAWRVSTKIRLPFSLRNIIDAGLLTEKEYEKLQD 170
QY 174 LG-----LP-HNTP-----WV-PWVWFANLSMKAYLGGRTRDTVLQSL 210
DB 171 INVSKAVILMLFVYVYTFKNLCETEPSGIRMLTPLHVVQOOLIDAEITAGR-GSVNYSV 229
QY 211 MNEVCTLTQCGOLYADWDISIPLYVTVVAVYSFFLACIGRQFLNPNKDYPGHMD 270
DB 230 TNLKAYRISPRLLYCHDWCVPLVYTVQVAALATYSYFFFLCGQDLNHDFFY---SLD 286
QY 271 LVVPVFTILQFLFYMGWLVKVAEQNLNPGEDDDDPETWIIDRLNQVLLSVGQHQMLP 330
DB 287 AFPLPLFTVQFLFFVGFVQGLMRPGLDLDLDFELSYLDRLNIVTSFTIVDSLQDDP 346
QY 331 P-MEEDMYW 338
DB 347 PKFEEDVFW 355
CC
RESULT 14
YAVK CAEL
ID YAVK CAEL STANDARD; PRT; 884 AA.
AC Q17851;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C0989.3 in chromosome IV.
GN C0989.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bradshaw H.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

```

```

CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; U50069; AAB37559.2; -.
CC DR WormPep; C0989.3; CE29571.
CC DR InterPro; IPR000615; Worm_fam_8.
CC DR Pfam; PF01062; DUF289; 2.
CC DR ProDom; PD002802; Worm_fam_8; 3.
CC KW Hypothetical protein; Transmembrane; Repeat.
CC FT TRANSMEM 265 285
CC FT TRANSMEM 398 418
CC SQ SEQUENCE 884 AA; 102851 MW; 8DB83F9699E8FALD CRC64;
CC
Query Match 18.5%; Score 546; DB 1; Length 884;
Best Local Similarity 32.6%; Pred. No. 2.3e-36;
Matches 125; Conservative 80; Mismatches 156; Indels 22; Gaps 10;
QY 1 MTITYNKVANRLGSSLLLCWRGSIYKLYGELFVIFLYSIRGLYRWVLSSDQQL 60
DB 1 MTISYTVDVATESYVGFVKVLPKWSGVKLIHRELFLMLVLYTVLAIYR-TLDEERKK 59
QY 61 LFEKALYCDSDYIQLP--ISFVLGFPVTLVSRWMSQYENLPWDLMLQVSSFVEGKD 118
DB 60 IFRS---NIEHFNPEPSILTFMLSPFTTVIVORWNNVFTMGFIENAAVVSFMK-NG 115
QY 119 EGRLLRRTLRVAILGQVLIRSIYSTVYKRPFTLHLVLAFGFMTHGEHKQLQKGLPH 178
DB 116 EDVRAQRTVIRVLVASQILVMSISIKALRRPNTVESIVTAGFLTKEESTIQNTLSY 175
QY 179 NTFVPMVWFANLSMKAYLGGRTRDTVLQSLMNEVCTLTQCGQLYADWDISIPLYTQ 238
DB 176 DSSCVPIRMAIQVLRHQYRSGNFFSHSVYRATWKEVSDPFETHLSRVKRVKVDWVPIAYPQ 235
QY 239 VTVVAVYSFFLACIGRQFLNPNKDYPGHMDLVVPVFTILQFLFYMGWLVKVAEQNLNPF 298
DB 236 VIFVAVLYEIVICAFKQYFDLDDDDARVYIHVYFFIVTFQFICLGMGLKVAEQNLNPL 295
QY 299 GEDDDDPETWIIDRLNQVLLSVGQHQNLPPMERDYMWNEAAPQPPYTAASARSR--- 355
DB 296 GEDDDDPEVNFLLDSNIYMT-----MSYNC-DVATDSYFNFFKVN-PFTKSTYKNNIF 347
QY 356 --RHSFMGSTFNISLKKEDELMW 376
DB 348 QILFRWKGSVK-SIWKE-LALW 368
CC
RESULT 15
YHS4 CAEL
ID YHS4 CAEL STANDARD; PRT; 602 AA.
AC O18303;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK849.4 in chromosome I.
GN ZK849.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]

```

Fri Jul 11 11:45:20 2003

us-09-622-964-29.rsp

RP REVISIONS.
RA Durbin R.; (OCT-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted 10/10/2000; to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82095; CAB05027.2; --
DR WormPep; ZK849.4; CE25696.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 2.
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 67607 MW; 458AB78802BD63A CRC64;

Query Match 17.1%; Score 502.5; DB 1; Length 602;
Best Local Similarity 29.9%; Pred. No. 4.7e-33;
Matches 135; Conservative 72; Mismatches 144; Indels 101; Gaps 12;

Qy 1 MTITYTKVANARLGSSLLLCWRGSIYKLLYGELFVIFLYYYSIRGLYRMVLSDDQL 60
Db 14 MTVTYNRAVSTESIHNFVSIMSYWHGSLVKSINKEIWIIVLYHLILFIRVMP----- 68

Qy 61 LFEKALYCDSYIQ-----LPISEVLGFYVTVLVSRWMSQYENLPMPLMIQVSS 112
Db 69 IP-GWADYCKKVIENWPHQDMTIPLEFLGFFVTVTDWRKAFQNIPIEICAFVAA 127

Qy 113 FVEG-----KDEGR----- 122
Db 128 AIPGRIRIINKONEVRAPGGQDLRDPAVEVNHVPQVGPAPAAALPMGAPERPAVAL 187

Qy 123 -----LLRRTLRVAILGQVILRSISTSVYKRFPTLHHLVLG 161
Db 188 PMGAPEGFAIQPSVDVKLTARTIIRYLVLSQILLFREISTYKRFVDLKLVDISK 247

Qy 162 FMTHGEHKLOK-LGLPH-NTFWPFWVWFANLSMKAYLGGRIRDTVLQSLMNEVCTLRT 219
Db 248 FLTDEELKILSKVKCDHYDSYFLPINWAFSILQE-----HKVKSNPENANVIRDWQV 303

Qy 220 QCGQLYAVDWISIPLVYTVTVAVYSPFLACLGROPINPNKDYPGHEMDLVVPVFTIL 279
Db 304 KLSLLRNGDFIPIPLAYPQAVFLAIRFYVLVCLFTROHLDMDDK---KSIDYFFPLMTSL 360

Qy 280 QFLYMGWGLKVAEQILNPPGDDDDPETNWIIDRLNQLVLSVD---GMHQNLPPMERDM 336
Db 361 QPFIYVGMWVAEILNPMGDDDDFELNNIIDKNLYGLAIVDTECGKH---PEIVKDT 417

Qy 337 YWNEAPOPPTATASARSRRHSFMGSTFNISL 368
Db 418 IGKDCLPFPQN-DDDNERNALVGSTKINL 448

Search completed: July 10, 2003, 12:24:47
Job time : 11.119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:21:29 ; Search time 30.163 Seconds
(without alignments)
3763.994 Million cell updates/sec

Title: US-09-622-964-29
Perfect score: 2944
Sequence: 1 MTITYTNKANRLGFSLSL.....EHAESYPYRDEAGTQVLYE 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB-seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1155	39.2	428	6 Q8WNR7	Q8wmx7 sus scrofa
2	1078.5	36.6	466	4 Q9NXP0	Q9nxd0 homo sapien
3	1069.5	36.3	465	11 Q8VCM0	Q8vcm0 mus musculu
4	961	32.6	721	5 Q9V3J6	Q9v3j6 drosophila
5	858.5	29.2	535	5 Q9VUM7	Q9vum7 drosophila
6	828.5	28.1	612	5 Q17206	Q17206 caenorhabdi
7	771.5	26.2	557	5 Q17205	Q17205 caenorhabdi
8	752	25.5	327	5 Q9VUM6	Q9vum6 drosophila
9	749.5	25.5	584	5 Q18303	Q18303 caenorhabdi
10	722.5	24.5	1447	5 Q16779	Q16779 caenorhabdi
11	714.5	24.3	731	5 Q9VRM4	Q9vrm4 drosophila
12	637	21.6	551	5 Q9NA59	Q9na59 caenorhabdi
13	616	20.9	512	5 Q62095	Q62095 caenorhabdi
14	559	19.0	551	5 Q965X4	Q965x4 caenorhabdi
15	503	17.1	400	5 Q17674	Q17674 caenorhabdi
16	495	16.8	444	5 Q17671	Q17671 caenorhabdi

17	168	5.7	72	4	Q9BR80	Q9br80 homo sapien
18	167.5	5.7	98	5	O01649	O01649 drosophila
19	117.5	4.0	516	16	Q8ZP99	Q8zpp9 salmonella
20	114.5	3.9	516	16	Q8Z772	Q8z772 salmonella
21	110.5	3.8	1139	4	Q9E278	Q9p278 homo sapien
22	109.5	3.7	1707	5	Q17652	Q17652 caenorhabdi
23	109.5	3.7	2606	12	Q36414	Q36414 alcelaphine
24	108.5	3.7	560	10	O64901	O64901 eschscholzi
25	108	3.7	1311	6	Q9BDH6	Q9bdh6 equus caball
26	102.5	3.5	331	16	Q9YVCO	Q9yvco anabaena sp
27	102.5	3.5	362	5	O17030	O17030 caenorhabdi
28	102.5	3.5	553	5	Q17414	Q17414 caenorhabdi
29	101	3.4	255	5	Q9NAE5	Q9nae5 caenorhabdi
30	101	3.4	503	8	Q8WKM3	Q8wkm3 hlyotelephi
31	100.5	3.4	532	10	Q9LINS	Q9lins arabidopsis
32	100	3.4	388	5	Q22454	Q22454 caenorhabdi
33	100	3.4	439	16	Q8XNR9	Q8xnr9 clostridium
34	99.5	3.4	685	16	Q9KUG5	Q9kug5 vibrio chol
35	99.5	3.4	854	10	Q9SXH8	Q9sxh8 brassica ol
36	99.5	3.4	1027	16	Q9SR88	Q9sr88 mycoplasma
37	98.5	3.3	518	3	O74283	O74283 coprinus ci
38	98	3.3	467	16	Q9XAR1	Q9xar1 streptomyce
39	97.5	3.3	533	17	Q96Z47	Q96z47 sulfobus
40	97.5	3.3	1349	5	Q9N5D3	Q9n5d3 caenorhabdi
41	97.5	3.3	1413	5	Q9NBD3	Q9nbd3 caenorhabdi
42	97	3.3	616	8	Q9G859	Q9g859 malawimonas
43	97	3.3	642	5	Q9V4V6	Q9v4v6 drosophila
44	97	3.3	856	5	Q963T4	Q963t4 drosophila
45	96.5	3.3	445	10	Q08700	Q08700 brassica na

ALIGNMENTS

RESULT 1

Q8WNR7	PRELIMINARY;	PRT;	428 AA.
ID	Q8WNR7		
AC	Q8WNR7;		
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Bestrophin (fragment)		
OS	Sus scrofa (pig)		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9823;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Marmorstein L.Y., McLaughlin P.J., Stanton B., Yan L., Crabb J.W.,		
RA	Marmorstein A.D.;		
RT	"Bestrophin interacts with the b-catalytic subunit of protein		
RT	phosphatase 2A";		
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY064707; AAL40882.1; -		
DR	InterPro; IPR000615; Worm_fam_8.		
DR	Pfam; PF01062; DUF289; 1_		
DR	ProDom; PD002802; Worm_fam_8; 1.		
FT	NON TER 1		
FT	NON TER 1		
SQ	SEQUENCE 428 AA; 48498 MW; 3E591B2BCD6383 CRC64;		
Query Match			
Best Local Similarity 55.0%; Pred. No 2.2e-91;			
Matches 241; Conservative 48; Mismatches 85; Indels 64; Gaps 12;			
QY	156	HLVLGAFMTGHGKLOKLGPHNTFWPVPWFANLSMKAYLGGRIRDTVLLQSLMNEVC	215
Db	3	HLVKGAFMTPESEKLEKLSLPHNSFWPVPWFANLSMKAYLGGRIRDPVLLQSLDMNN	62
QY	216	TLRTCGQLYADWISLIPVTVTVTVVAVYSFFLACLGQFLNPKNVYFGHEMDLVVPV	275
Db	63	TLRTCGQLYADWISLIPVTVTVTVVAVYSFFLACLGQFLNPKNVYFGHEMDLVVPV	122
QY	276	FTILOFLYMGWLKVAEQILNPFGEDDDDFTNWIIDRLNQVLSLLSDVCGHQLPMPERD	335

123 FTFLQPFYAGLWVAEQLINPFGEDEDDFETNWIIVDRSLQVLSAYDEMHDLPPMERD 182
336 MYWNEAAPQPPYTAASARSRRHSFMGSTFNISLKEDLELWSKEADTDKESGYSSTIG 395
183 MYWNPDPHPYTAASQAQRSPSPFGSTFNISLKGDEMEPOPEE-----EAHTGILG 236
396 CFLGLQPKNYHLPLKDLTKLLCSKNPLLEGQCKDANOKN-----QKD--VWFKF 443
237 HFLGLQSSDHQPPRTNKTLLWPKK---EGHFEGHPKNLARGARLSDSDQEDSKANREG 293
444 GLDPLKVPKFRGSGCGPOAPSSH-----PTQSQAP-----SSSDTGDGPST 487
294 G---FSAALCGRPGYHSAPOTPLGHTPMVPPPEAPLGLRRVSGIDEAAKQDLOPAT 350
488 -----DYQEIChMKKKTVEPNL-NIPESPTEHQORRLDQSTWIOA 528
351 PSYKSPFELIPESAEAPLQGGSHVRKTVFENLADLSEAP-EHLKEPNLEP-PMGIIHA 408
529 LMKHEAESY---PYRDEA 543
409 ILKDRDPYWALENRDEA 426

RESULT 2

Q9NXP0 PRELIMINARY; PRT; 466 AA.
AC Q9NXP0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE CDNA PLJ20132 fis, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK00139; BAA90970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

Query Match 36.6%; Score 1078.5; DB 4; Length 466;
Best Local Similarity 60.1%; Pred. No. 9.9e-85;
Matches 199; Conservative 57; Mismatches 74; Indels 1; Gaps 1;

QY 45 SIRGLRMVLSDDQQLFEKALYCDSYIQLIPISVLGFGYVTLVSRWWSQYENLPWD 104
DB 2 ALSAAYRFLVTEGQKRYFEKLVICDYQYASLIPVSVLGFYVTLVWNRWWSQYENLPWD 61
QY 105 RLMIQVSSFFVEGKDEEGRLLRRLIRYAILGOVLIRLSISTSVYKRPPTLHLVLGFM 164
DB 62 ALMCVAGTVHGDRGLRRLRLMYAGLSAVLILRSVSTAVFKRPTTDHVVAGFM 121
QY 165 HGEHKQLQKLGPHNTFWPVMWPNFANLSKAYLGGRIRDTVLQSLMNEVCTLRQCGL 224
DB 122 REERKKFENLSSYNKYVPCWVFNLSAAQARREGIRDSNLSKLLLELVNFRKCGML 181
QY 225 YADWISIPLVYTVVTVVAVYSFELACIGROFLPNKDYPGHEMDLVVPVFTILOFLFY 284
DB 182 FHYDWSIPLVYTVVTVVAVYSFELACIGROFLPNKDYPGHEMDLVVPVFTILOFLFY 241
QY 285 MGWLKVAEQILNPFGEDEDDFETNWIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 344
DB 242 AGWLKVAEQILNPFGEDEDDFETNFIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 301

QY 345 PPYTAASA-RSRRHSFMGSTFNISLKEDLE 374
DB 302 APYTAATVFLQRPSPFGSTFDITLAKEDMQ 332

RESULT 3

Q8VCM0 PRELIMINARY; PRT; 465 AA.
AC Q8VCM0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ20132.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019528; AAH19528.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51998 MW; E7145E2F893B03F6 CRC64;

Query Match 36.3%; Score 1069.5; DB 11; Length 465;
Best Local Similarity 58.7%; Pred. No. 5.9e-84;
Matches 199; Conservative 59; Mismatches 80; Indels 1; Gaps 1;

QY 45 SIRGLRMVLSDDQQLFEKALYCDSYIQLIPISVLGFGYVTLVSRWWSQYENLPWD 104
DB 2 ALSAAYRFLVTEGQKRYFEKLVICDYQYASLIPVSVLGFYVTLVWNRWWSQYENLPWD 61
QY 105 RLMIQVSSFFVEGKDEEGRLLRRLIRYAILGOVLIRLSISTSVYKRPPTLHLVLGFM 164
DB 62 ALMCVAGTVHGDRGLRRLRLMYAGLSAVLILRSVSTAVFKRPTTDHVVAGFM 121
QY 165 HGEHKQLQKLGPHNTFWPVMWPNFANLSKAYLGGRIRDTVLQSLMNEVCTLRQCGL 224
DB 122 REERKKFENLSSYNKYVPCWVFNLSAAQARREGIRDSNLSKLLLELVNFRKCGML 181
QY 225 YADWISIPLVYTVVTVVAVYSFELACIGROFLPNKDYPGHEMDLVVPVFTILOFLFY 284
DB 182 FHYDWSIPLVYTVVTVVAVYSFELACIGROFLPNKDYPGHEMDLVVPVFTILOFLFY 241
QY 285 MGWLKVAEQILNPFGEDEDDFETNWIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 344
DB 242 AGWLKVAEQILNPFGEDEDDFETNFIIDRLNQLSVLSVDMGHQNLPPMERDMYWNEAAPQ 301
QY 345 PPYTAASA-RSRRHSFMGSTFNISLKEDLELWSKEAD 382
DB 302 APYTAATAFLQRPSPFGSTFDITLAKEDMQFORLDGVD 340

RESULT 4

Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6264 protein (LP05915P).
GN BEST OR CG6264.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```
Query Match      : 32.6%; Score 961; DB 5; Length 721;
Best Local Similarity 50.4%; Pred.No. 2.5e-74;
Matches 191; Conservative 70; Mismatches 108; Indels 10; Gaps .

Oy 1 MTITYNKVANAR-LGSFSLLCWRGSIYKLYGEFLVFIPLYYSIRGLYRWLSSDQ 59
    ||||| : | | | | | | | | | | | | | | | | | | | | | |
Db 1 MTITYGEVATCRGFCFKLLLRWRSIYKLVWLDFLLAFIYYAINNVYRFGLNPAOK 60
```

```
Query Match      : 32.6%; Score 961; DB 5; Length 721;
Best Local Similarity 50.4%; Pred.No. 2.5e-74;
Matches 191; Conservative 70; Mismatches 108; Indels 10; Gaps .

Oy 1 MTITYNKVANAR-LGSFSLLCWRGSIYKLYGEFLVFIFLYYSIRGLYRWLSSDQ 59
    ||||| : | | | | | | | | | | | | | | | | | | | | | |
Db 1 MTITYGEVATCRGFCFKLLLRWRSIYKLVWLDFLLAFIYYAINNVYRFGLNPAOK 60
```

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RT Science 287:2185-2195(2000).
 DR EMBL; AB003531; AAF49648.1; --
 DR FlyBase; FBgn0036492; CG12327.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 535 AA; 62710 MW; 7C9333A0E5068DF2P CRC64;

Query Match 29.2%; Score 858.5; DB 5; Length 535;
 Best Local Similarity 44.0%; Pred. No. 1.2e-65;
 Matches 165; Conservative 74; Mismatches 129; Indels 7; Gaps 4;

QY 1 MTITYTNKVAN-ARLGSFSLLCWRSYKLLYGBFLVFIPLYISIRGLYRMVLSDDQ 59
 DB 1 MTVSYTAEVATCSHFCKWLLMRWRASYIKIIVDLLAFLSCFYFMVYRYALRDVDK 60

QY 60 LPEKALYCDYIQLIPISFVLGVVTVLVSRWMSQYENLPWPDRLMTQVSVFVGKDE 119
 DB 61 PVFEDIVMYCHSYNLIPISFVLGVVTVLVSRWMSQYENLPWPDRLMTQVSVFVGKDE 120

QY 120 EGRLLRRLIRYAILQGVLLRSISVYKRPPTLHLVLGAGFTHGHEKQLKGL--- 176
 DB 121 HGRLLRRLIRYAILQGVLLRSISVYKRPPTLHLVLGAGFTHGHEKQLKGL--- 180

QY 177 PNTFWPVPWVPANLSKAYLGRIRDTVLQSLMNEVCTLRQCQQLYAYDWISIPLVY 236
 DB 181 KHPKYWPIVWASIVTRAKRGIWDDPSLKSMDLANKFRAGCNWLHYDIISVPLY 240

QY 237 TQVTVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLFYMGWLKVAEQLIN 296
 DB 241 TQVTVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLFYMGWLKVAEQLIN 300

QY 297 PFGEDDDDFETNWIIDRNLOVSLLSVDGMHONLPWMDRMYWNEAAP-QPPYTAASARS 355
 DB 301 PFGEDDDDFELNWLIDRNLOVSLLSVDGMHONLPWMDRMYWNEAAP-QPPYTAASARS 360

QY 356 RHSPFMGSTFNISLKK 370
 DB 361 HPE--ASTARLGIK 373

RESULT 6
 O17206 PRELIMINARY; PRT; 612 AA.
 AC O17206;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C01B12.3 protein.
 GN C01B12.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kerahaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaison N., Smith A., Sonnhammer E., Staden R., Sulston J.,

RA Thierzy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Scheet P., Maggi L.;
 RT "The sequence of *C. elegans* cosmid C01B12.1";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025458; AAB70976.1; --
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 612 AA; 71031 MW; DFB843916541DD44 CRC64;

Query Match 28.1%; Score 828.5; DB 5; Length 612;
 Best Local Similarity 38.3%; Pred. No. 5.6e-63;
 Matches 160; Conservative 86; Mismatches 145; Indels 27; Gaps 4;

QY 1 MTITYTNKVANARLGSFSLLCWRSYKLLYGBFLVFIPLYISIRGLYRMVLSDDQ 60
 DB 1 MTVTSLDVASSFPCLYKLLFRWKSINKSWAELVWVWCLYAVLSVIYRCLLTKMQR 60

QY 61 LPEKALYCDYIQLIPISFVLGVVTVLVSRWMSQYENLPWPDRLMTQVSVFVGKDE 120
 DB 61 TFEDLCIFDFTSYNPIPTFMLGFFVSAVFTRWQIFDNIGWIDTPTCLMTQVKGTER 120

QY 121 GRLLRRLIRYAILQGVLLRSISVYKRPPTLHLVLGAGFTHGHEKQLKGLPHNT 180
 DB 121 AKCVRENCIRYSILTOAMVYRDVAASVRKRPFTFNHLVLTAGLMTKEMAEFESIPSHAK 180

QY 181 FVVPWVPWVPANLSKAYLGRIRDTVLQSLMNEVCTLRQCQQLYAYDWISIPLVY 240
 DB 181 YQCPMHLFSMITLARDEGMISDIIYVDLMEKQRFVNVILSLTFDWPVPLVYTVQV 240

QY 241 TVAVYSFFLACIGROFLNPNKDYPGHENDLVVVPVFTILOFLFYMGWLKVAEQLIN 296
 DB 241 HLAVRSYFLIALFGROYLHPESNRLNDFK-QTIDLVVPMISLLQFIFFIGMKVAVELN 299

QY 297 PFGEDDDDFETNWIIDRNLOVSLLSVDGMHONLPWMDRMYWNEAAPQPPYTAASARS 356
 DB 300 PFGEDDDDFECNWLIDRNLOVSLLSVDGMHONLPWMDRMYWNEAAPQPPYTAASARS 359

QY 357 HSPFMGSTFNISLKKED-----LELWSKBEADTD-----KESGYS 392
 DB 360 NQVGS CADMPTEEBEPFVPRPRRLTSLRSHWDGDMEDTDVVPVGLKHTRDNSNYAS 417

RESULT 7
 O17205 PRELIMINARY; PRT; 557 AA.
 AC O17205;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C01B12.5 protein.
 GN C01B12.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RESULT	8
QSVUM6	PRELIMINARY; PRT; 327 AA.
ID	
AC	Q9VUM6;
DT	01-MAY-2000 (TRENBLrel. 13, Created)
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE	CG7259 protein.
GN	CG7259.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BERKELEY;
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George K.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA	Ballew R.M., Basu A., Bakendate J., Bayraktaroglu L., Beasley E.M.,
RA	Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Buck J., Brokstein P., Brotter P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferriaz C., Fleischmann W.,
RA	Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA	Hoetin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kammel B.S., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., MCPerson D.,
RA	Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-P., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster.";
RL	Science 287:2185-2195(2000).
RE	EMBL; AE003531; AF49649.1; ..
DR	FlyBase; FBgn0036491; CG7259.
DR	InterPro; IPR000615; Worm_fam_8.
DR	Pfam; PF01062; DUF289; 1.
DR	ProDom; PD002802; Worm_fam_8; 1.
SQ	SEQUENCE 327 AA; 30594 MW; BE7CE3BC05370F46 CRC64;
Query Match	25.5%; Score 752; DB 5; Length 327;
Best Local Similarity	47.7%; Pred. No. 9.9e-57;
Matches 143; Conservative	54; Mismatches 97; Indels 6; Gaps 3
QY	68 YCDSYIQLPISFVLGFFVTTLVVSRWMSQVENLPWDELMTQVSSFEVGKDEGLLRRT 127
Db	3 YCENGLSPLSFVLGFFVRIVNNRWMDQYTTTPDPGIAITLISTSHGSDDRAWMERT 62
QY	128 LRIRAILGVQLIIRSTISTSYVKFPPTLHLVLACFMTHGEHKOLQKLGHNTPE---WVP 184

Db 63 ILRYVCLQVIVFTMISPRVREPPYITQIIIEAGFLLENEKKIIEAMDOAFSPYKHPMP 122
Qy 185 WYVFNALSKAYLGGRIIDTVLQSLMNEVCTLTCTGGQGLYAYDVWISILVYTOVTVAV 244
Db 123 IYVAAISVNRARENKIRDDYAVKTIIDELNQLRGNGCPFLYYDWSVPLVYTVAT 182
Qy 245 YSFLACLIGROFLNPNKQYPGHEMDLVVFTILOFLYMGWLVKVAEOLINPFGDDDD 304
Db 183 YSFLFSVLGQQWNESSHSDGTRIRRWPFILVQLQFFYMGWLVKVAETLINPFGDDDD 242
Qy 305 FETNWIIDRLNQLVSLSDVGMHQLNPMERDMYNEAAP-OPYTAAASRRHSFWSGT 363
Db 243 FELNWIIDRLNQLVSLSDVGMHQLNPMERDMYNEAAP-OPYTAAASRRHSFWSGT 300

RESULT 9
Q18303 PRELIMINARY; PRT; 584 AA.
AC Q18303, 1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE C29F4.2 protein.
GN C29F4.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z68335; CAA92730.1; -
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 584 AA; 68252 MW; 86A86A4EEB21BCC CRC64;

Query Match 25.5%; Score 749.5; DB 5; Length 584;
Best Local Similarity 39.4%; Pred. No. 3.5e-56;
Matches 139; Conservative 81; Mismatches 132; Indels 1; Gaps 1;

Qy 20 LLLCWGSIYKLYGFEFLVIFLYYSIRGLYRMVLSQQLLFEKALYCDSYIQLIPIS 79
Db 64 LIFRWKGLSKAYILDLIVWCFCYAFISVIVYALDRSQDTFFRPMQFCNRLDYIPIN 123
Qy 80 FVLGFFVTLVSVRWMSQYENLPWDLMIQVSSFVEGKDEGRLLRRLTRIYAILGQVLI 139
Db 124 FVLGFFVTLVSVRWMSQYENLPWDLMIQVSSFVEGKDEGRLLRRLTRIYAILGQVLI 183
Qy 140 LRSISTSVKRPPTLHLVLGFMTHGEKQKQLGLPHNTFWVWVFNLSMKAYLGG 199
Db 184 FRDIHIGVKKRPPTLHLVLGFMTHGEKQKQLGLPHNTFWVWVFNLSMKAYLGG 243
Qy 200 RIRDTVLQSLMNEVCTLTCTGGQGLYAYDVWISILVYTOVTVAV 259
Db 244 RIESAYTQNAIAEERTFRASGLSLWTYDWPILMYPOLVFWAHCYLVCLVSRQFVI 303
Qy 260 PNKOYFGHEMDLVVFTILOFLYMGWLVKVAEOLINPFGDDDDDFETNWIIDRLNQLVSL 319
Db 304 NSDAVNTTIDLGVPMTIIEIFYMGWLVKVAEOLINPFGDDDDDFETNWIIDRLNQLVSL 363
Qy 320 LSVDCGHQNLPMERDMYNEAAPOPYTAAASRRHSFWSGTENISLKKED 372
Db 364 GIVDSDHDDGPILEKDMFWNTV-SPLYSSAAQRNWNFFGSAATNADAQIFD 415

RESULT 10
O16779 PRELIMINARY; PRT; 1447 AA.
AC O16779; O16781; PRT; 1447 AA.
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 164.2 kDa protein.
GN T21D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Moessner J.;
RL "The sequence of C. elegans cosmid T21D12.";
RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016687; AAK72064.1; -
DR HSP; F56276; ITLK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_tyr.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR Pfam; PF00047; Ig; 3.
DR Pfam; PF00560; LRR; 12.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR ProDom; PD002802; Worm_fam_8; 1.
DR SMART; SM00409; Ig; 3.
DR SMART; SM00408; Ig_c2; 2.
DR SMART; SM00410; Ig_like; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_tyr; 13.
KW Hypothetical protein, immunoglobulin domain.
SQ SEQUENCE 1447 AA; 164168 MW; 4BC1F38E9E8E5B5 CRC64;

Query Match 24.5%; Score 722.5; DB 5; Length 1447;
Best Local Similarity 36.4%; Pred. No. 2.5e-53;
Matches 160; Conservative 87; Mismatches 168; Indels 25; Gaps 9;

Qy 1 MTITVTKVAVARLGSFSSLLCWGSIYKLYGFEFLVIFLYYSIRGLYR--WVLSDDQ 58
Db 879 MTSVNSDVSSTSLIAFLRLQLRWGWSVMSKRELIFPSCFALITSIYTNVFLSDEQ 938
Qy 59 QLLFEKALYCDSYIQLIPISFVLGFFVTLVSVRWMSQYENLPWDLMIQVSSFVEGKD 118

```

Db 939 RVFMDNFAALFQDTLNYILPTFMLGPFVITIIIVGRWMDIFLNLGWDVNTALLIATYIRGD 998
Qy 119 BEGRLLRRLTLYAILGQVLLIARSISTSVYKRPPTLHLVLVLAGFWTHGKHQLOKLG-L-P 177
Db 999 EKSRIMRTALRYLVLTQVLIIFRDISLVYKRPPTLETIVVAGSFMLSESEKRYAALSLGK 1058
Qy 178 HNTFVVPVWVFWANLKMAYLGRIRDTVLLQSLMNEVCTLRQCQGLYAYDWISIPLVYT 237
Db 1059 YPKYFLPQWCFSLYDARAQKIGADVWMLNLIKSVGDFRGLGQLLNDFWPIPLVTP 1118
Qy 238 QVTVVAVVSFFLACILGRQL-----NPNKDPYGHMDLVVP-VFTILOFLFYMGMLKVA 291
Db 1119 QVFLAVIYFPLCLILARQSVLIDGKPKNSVVPF---VFLMTSLQFVYVGMKVA 1175
Qy 292 EQLINPFGEDDDDFETNIIIDRNLOVLSLVGDMQNLPPMRDMYWNAAPOPPYTAAS 351
Db 1176 ESLMPLGEDDDDFECNVLDRNLAVGLAIVDS-SEAAPHYKDVFWGLAIADPLYSDT 1234
Qy 352 ARSRHRSFWGTFNLSLKEDLELWKEADTDKESGVSYSTIGC-----FLGLQPKN 404
Db 1235 ANIHINPOGSA--ATYETQNEIITRPHVDINDMEDGDDVEGCNPROLSRFVSVVN 1292
Qy 405 YHLPKDKLTKLCSKNPL 424
Db 1293 RNC---ESRTSLSSRNPTI 1309

```

RESULT 11

```

Q9VRW4 ID Q9VRW4 PRELIMINARY; PRT; 731 AA.
AC AC Eukaryota; Metazoa; Arthropoda; Insecta;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG10173 protein.
GN CG10173
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Bau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Esler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nueskern D.R., Pacle J.M.M.,
RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sien-Kiam I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Voh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AB003563; AAF50668.1; -.
DR FlyBase; FBgn0035696; CG10173.
DR InterPro; IPR000815; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 731 AA; 80654 MW; FF4D625528826C6D CRC64;

Query Match 24.3%; Score 714.5; DB 5; Length 731;
Best Local Similarity 46.7%; Fred. No. 5e-53;
Matches 142; Conservative 50; Mismatches 103; Indels 9; Gaps 4;

Qy 78 ISFVLGFTVTLVSRWWSQYENLPMPLMIQVSVFVEG-----KDEGRLLRRLTIRYA 132
Db 1 MSFVLGFTVTLVSRWWSQYENLPMPLMIQVSVFVEG-----KDEGRLLRRLTIRYA 60

Qy 133 ILGQVLILRSISTSVYKRPPTLHLVLVLAGFWTHGKHQLOKLG--LPHTFVVPVWVFAN 190
Db 61 VLAVITLQIRSLKRRFPPTQHLVDAGLHSEMKIFEALNOKSPHMYWPLVWATN 120

Qy 191 LSMKAYLGRIRDTVLLQSLMNEVCTLRQCQGLYAYDWISIPLVYTQVTVVAVSFPLA 250
Db 121 IINRARKDGLIASDHIVQTLVLSLDIRRLGGLGYDTCVPLVYTQVTVVAVSFPLA 180

Qy 251 CLIGRQFLNPKDYPGHE-MDLVVPVFTILOFLFYMGMLKVAEQLINPFGEDDDDFETNW 309
Db 181 ALLGRQMLPNVLDGRSDPDLFPFLFTVLQFVYVGMKVAEVLNPFGEDDDDIELNW 240

Qy 310 IIDRNLOVLSLVGDMQNLPPMRDMYWNAAPOPPYTAASARSRHRSFWGTFNLSL 368
Db 241 LIDRHIKAAYMIVDEMHEHPPELLRDQWECVVPKDLFTYVASEHYRDEPKGSAEKYKV 300

Qy 369 KKED 372
Db 301 KKED 304

RESULT 12
Q9NA59 ID Q9NA59 PRELIMINARY; PRT; 551 AA.
AC Q9NA59;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y73F8A.11 protein.
GN Y73F8A.11
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 1
RP SEQUENCE FROM N.A.
RC MEDLINE=99069613; PubMed=9851916;
RA Matthews L.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; AL132862; CAB60536.1; -.
DR InterPro; IPR000615; Worm_fam_8.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:21:29 ; Search time 32.0242 Seconds
(without alignments)
3763.954 Million cell updates/sec

Title: US-09-622-964-3
Perfect score: 3120
Sequence: 1 MTITYTSQVANARLGSFRL.....TLKDHMPYALENDEAHS 585

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1701	54.5	428	6	Q8WNR7	Q8wmr7 sus scrofa
2	1137.5	36.5	465	11	Q8VCM0	Q8vcm0 mus musculu
3	1133.5	36.3	466	4	Q9NXP0	Q9nxp0 homo sapien
4	1009	32.3	721	5	Q9V3J6	Q9v3j6 drosophila
5	886	28.4	535	5	Q9VUM7	Q9vum7 drosophila
6	814	26.1	612	5	O17206	O17206 caenorhabdi
7	770.5	24.7	557	5	O17205	O17205 caenorhabdi
8	758	24.3	584	5	O18303	O18303 caenorhabdi
9	742	23.8	327	5	Q9VUM6	Q9vum6 drosophila
10	717.5	23.0	1447	5	O16779	O16779 caenorhabdi
11	694.5	22.3	731	5	Q9VRW4	Q9vrw4 drosophila
12	622.5	20.0	551	5	Q9NA59	Q9na59 caenorhabdi
13	616.5	19.8	512	5	O62095	O62095 caenorhabdi
14	554	17.8	551	5	Q965X4	Q965x4 caenorhabdi
15	495	15.9	400	5	O17674	O17674 caenorhabdi
16	455	14.6	444	5	O17671	O17671 caenorhabdi

17	175	5.6	72	4	Q9BR80	Q9br80 homo sapien
18	159.5	5.1	98	5	O01649	O01649 drosophila
19	117	3.8	516	16	Q8ZP99	Q8zpp99 salmonella
20	115	3.7	516	16	Q8Z772	Q8z772 salmonella
21	113.5	3.6	700	10	Q94874	Q948j4 oryza sativ
22	112	3.6	710	10	Q948J6	Q948j6 oryza sativ
23	112	3.6	1256	5	Q9W238	Q9w238 drosophila
24	111.5	3.6	798	4	Q9BWG7	Q9bwg7 homo sapien
25	109	3.5	1051	10	Q9SE97	Q9se97 arabidopsis
26	107.5	3.4	1222	5	Q9TXV2	Q9txv2 caenorhabdi
27	107	3.4	2606	12	O36414	O36414 alcelaphine
28	106.5	3.4	416	2	Q9F2A1	Q9f2a1 chromatium
29	105	3.4	683	11	Q99MT4	Q99mt4 mus musculu
30	105	3.4	786	11	Q99MT3	Q99mt3 mus musculu
31	105	3.4	1707	5	Q17652	Q17652 caenorhabdi
32	104.5	3.3	694	5	Q8T4D2	Q8t4d2 drosophila
33	104.5	3.3	785	10	Q9SA37	Q9sa37 arabidopsis
34	104	3.3	485	3	O13920	O13920 schizosacch
35	104	3.3	758	10	Q8S9K3	Q8s9k3 arabidopsis
36	104	3.3	886	3	Q9P7N3	Q9p7n3 schizosacch
37	103.5	3.3	356	4	Q9NUU6	Q9nuu6 homo sapien
38	103.5	3.3	741	16	Q8X7V5	Q8x7v5 escherichia
39	103	3.3	342	12	O90387	O90387 kaposi's sa
40	103	3.3	681	10	Q9SX10	Q9sxi10 oryza sativ
41	103	3.3	758	10	Q9FN73	Q9fn73 arabidopsis
42	103	3.3	1359	5	Q8T6H8	Q8t6h8 dictyosteli
43	102.5	3.3	491	5	Q960V9	Q960v9 drosophila
44	102.5	3.3	569	5	O3VL56	O3vl56 drosophila
45	102.5	3.3	1002	3	O13849	O13849 schizosacch

ALIGNMENTS

RESULT 1
Q8WNR7 ID Q8WNR7 PRELIMINARY; PRT; 428 AA.
AC Q8WNR7; ID Q8WNR7; TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bestrophin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_taxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Marmorstein L.Y., McLaughlin P.J., Stanton B., Yan L., Crabb J.W.,
RA Marmorstein A.D.;
RT "Bestrophin interacts with the b-catalytic subunit of protein
RT phosphatase 2A".
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY064707; AAL40882.1; -
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
FT NON TER 1
SQ SEQUENCE 428 AA; 48498 MW; 3B6591B2BCDB6383 CRC64;
Query Match 54.5%; Score 1701; DB 6; Length 428;
Best Local Similarity 74.8%; Pred. No. 7.4e-135;
Matches 324; Conservative 29; Mismatches 72; Indels 8; Gaps 4;
Qy 155 QHLVQAGFMTPEAKHLEKLSLPHNMFWVFWFWANLSMKAWLGGRTDPILLQSLNEM 214
Db 2 QHLVQAGFMTPEAKHLEKLSLPHNMFWVFWFWANLSMKAWLGGRTDPILLQSLNEM 61
Qy 215 NTLRTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHDLVVP 274
Db 62 NTLRTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHDLVVP 121
Qy 275 VFTFLQFFVYVWGLKVAEQLINPFGEDEDDDDPFTNWIVDRLNQLVSLLAVDENHQLPRMEP 334

```

Db 122 LFTFLQPPFYAGWLVAEQINLPPGDDDDFTNIVDRSLQVSLAVDEMHDLPMPER 181
Qy 335 DWYWNKPPQPPYTAASQAFPRASFMGSTFNISLNKEEMEFQNOEDEDAHAGIIGREL 394
Db 182 DWYWNKPPQPPYTAASQAFPRASFMGSTFNISLNKEEMEFQNOEDEDAHAGIIGREL 239
Qy 395 GLOSHDHPHPPRANSRTKLLWPKRESLLHGLHPKHNKAAKQNVRGQEDNKAWKLVADAPK 454
Db 240 GLOSHDHPHPPRANSRTKLLWPKRESLLHGLHPKHNKAAKQNVRGQEDNKAWKLVADAPK 296
Qy 455 SCPLYQRYGYVSAPOTPLSPMPFPLPSAPSKLHSTVGID--TKOKSLKTVSSGAKKS 512
Db 297 SAALCGRPGYSAPOTPLSGHTPMVFPFPEASPLGLRRVSGIDEAAKQSLQATPSIRKS 356
Qy 513 FELLSESCALMEHPEVSVORRKTVEFNLTDMPEIPENHLKEPQSPNTNIHTLKDHMD 572
Db 357 FELLSESCALMEHPEVSVORRKTVEFNLTDMPEIPENHLKEPQSPNTNIHTLKDHMD 415
Qy 573 PYWALENDRDEAHS 585
Db 416 PYWALENDRDEAHS 428

RESULT 2
Q8VCM0 PRELIMINARY; PRT; 465 AA.
AC Q8VCM0;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DE Similar to hypothetical protein FLJ20132.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019528; AAH19528.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51998 MW; E7145E2F893B03F6 CRC64;

```

```

Query Match 36.5%; Score 1137.5; DB 11; Length 465;
Best Local Similarity 51.9%; Pred. No. 2.5e-87;
Matches 242; Conservative 56; Mismatches 127; Indels 41; Gaps 10;

Qy 50 YRLALTEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNOYENLPWDLMSL 109
Db 7 YRLALTEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNOYENLPWDLMSL 66
Qy 110 VSGFVEGKDEQSLRLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDDRGLRYRRLTMRVYAGLSAVLILRSVSTAVYKRPSPSAQHLVQAGFMTREERK 126
Qy 170 QLEKLSLPHNMFVWVWVFNANLAKWLGGRIRDPILLQSLNEMNTLTQCGHLYAYDW 229
Db 127 KFNENLSSYNKYWVFCWFSNLAQAARREGIRDSNLSALKLLLELNVPFGKGMFLHYDW 186
Qy 230 ISPIVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLQFFYVGLWK 289
Db 187 ISPIVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLQFFYVGLWK 246
Qy 290 VAEQLINPFGEDDDFTNIVDRNLQVSLAVDEMHDLPMEPDYWNKPEPPPYTA 349
Db 247 VAEQLINPFGEDDDFTNIVDRNLQVSLAVDEMHDLPMEPDYWNKPEPPPYTA 306
Qy 350 ASA-QFRASFMGSTFNISLNKEEMEFQNOEDEDAHAGIIGRELQSHDHP 403
Db 247 VAEQLINPFGEDDDFTNIVDRNLQVSLAVDEMHDLPMEPDYWNKPEPPPYTA 306
Qy 350 ASA-QFRASFMGSTFNISLNKEEMEFQNOEDEDAHAGIIGRELQSHDHP 403

```

```

Db 307 ATAFLLQPPSQSGSFDIALAKEDMQFO--RLDQVGDPLGEVHGFLOELL-----P 356
Qy 404 PRANSRTKLLWPKRESLLHGLHPKHNKAAKQNVRGQEDNKAWKLVADAPKSGP 457
Db 357 AGAGSVGVL--GRRLLSLRR---KNSCVSEASTASCAGAADGGVVECGGDDPLLDPS 411
Qy 458 LYQ---RGGYSAPOTPL-SPTMPF-----PLEPSAPSKL 489
Db 412 LREPELEPPACPEPPAPIPGTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 457

RESULT 3
Q9NXPO PRELIMINARY; PRT; 466 AA.
AC Q9NXPO;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE CDNA FLJ20132 f18, clone COL06441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000139; BAA90970.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 466 AA; 52008 MW; BB9608363A09C2EC CRC64;

```

```

Query Match 36.3%; Score 1133.5; DB 4; Length 466;
Best Local Similarity 63.7%; Pred. No. 5.4e-87;
Matches 209; Conservative 50; Mismatches 68; Indels 1; Gaps 1;

Qy 50 YRLALTEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNOYENLPWDLMSL 109
Db 7 YRLALTEQQLMFEKLTLYCDSYIOLIPISFVLGFYVTLVTRWNOYENLPWDLMSL 66
Qy 110 VSGFVEGKDEQSLRLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHK 169
Db 67 VAGTVHGRDDRGLRYRRLTMRVYAGLSAVLILRSVSTAVYKRPSPSAQHLVQAGFMTREERK 126
Qy 170 QLEKLSLPHNMFVWVWVFNANLAKWLGGRIRDPILLQSLNEMNTLTQCGHLYAYDW 229
Db 127 KFNENLSSYNKYWVFCWFSNLAQAARREGIRDSNLSALKLLLELNVPFGKGMFLHYDW 186
Qy 230 ISPIVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLQFFYVGLWK 289
Db 187 ISPIVYTVQVTVAVYSFELTCLVGRQFLNPAKAYPGHGLDLVVPVFTFLQFFYVGLWK 246
Qy 290 VAEQLINPFGEDDDFTNIVDRNLQVSLAVDEMHDLPMEPDYWNKPEPPPYTA 349
Db 247 VAEQLINPFGEDDDFTNIVDRNLQVSLAVDEMHDLPMEPDYWNKPEPPPYTA 306
Qy 350 ASA-QFRASFMGSTFNISLNKEEMEFQ 376
Db 307 ATVFQLRQPSFGSFTDITLAKEDMQFO 334

```

```

RESULT 4
Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

```


01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
CG6264 protein (LP05915P).
BEST OR CG6264.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Cherry J.M., Cavley S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisic A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Taveanli B.C., Pappu K., Mardon G.;
RT "deest" a Drosophila Homolog of Human Bestrophin.";
RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Faragaa V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003686; AAF54503.1; -
DR EMBL: AF218817; AAF32327.1; -
DR EMBL: AY061546; AAL29094.1; -
DR FlyBase; FBgn0040238; best.
DR InterPro; IPR000615; Wormfam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Wormfam_8; 1.
SQ SEQUENCE 721 AA; 79642 MW; B37CD96DE2F15B15 CRC64;

Query Match 32.3%; Score 1009; DB 5; Length 721;
Best Local Similarity 36.7%; Pred. No. 3.2e-76;
Matches 250; Conservative 83; Mismatches 216; Indels 132; Gaps 18;

QY 1 MTITYTSQVANAR-LGSFSLRLCWRGSIYKLLYGEEFLIFLLCYIIIRFYRLALTBEQ 59
Db 1 MTITYTGEVATCRGFCFLKLLLRWGSYKLVWLDLAFLLIYAINMVYRFLNPAQK 60
QY 60 LMPEKLTLYCDSYIQLIPISFVLGFFYTVLVTRWNOYENLPWDRMLSLVSGFVEGDE 119
Db 61 ETEFAIVQYCDYRELIPLSFVLGFFYTVLVTRWNOYTSIPWDPPIAFVSSNVHGOE 120
QY 120 QSRLLRRLRYANLGNVLLRSYSTAVYKFPASQHLVQAGTWPABHKOELK--SLP 177
Db 121 RGRMRETIIRYVCLTULVLANSPVKKFGPGLNLDVEAGLNDNEKTIETMNAKFP 180
QY 178 H-NMFWPVPWFANLSMKAMLGGRIRDPILILQSLNEMNTLRTQGHLYAYDWISIPLY 236
Db 181 RPSKHNLPVWAAIIITRKEGRIRDDFAVKTIIDELNFKRGCGCLLSYDTISVPLVY 240
QY 237 TQVTVAVYSFPLTCLVGRQ-----FLNPAKAYPGHELDLVVPTFLQFFVY 285
Db 241 TQVTVAVYSFPLTCLVGRQ-----KVDLYPFVFTTLOFFVY 292
QY 286 GNLKVAEQLINPFGEDDDDDFETNIVDENLQVSLAVDEMHDLPMBEPDMYWNKPEP-Q 344
Db 293 GNLKVAESLINPFGEDDDDDFVNMVDRNLQVSLVIVDEMHHDPHLLKQDWVDFPNE 352
QY 345 PPYTAASQPRFRASFMGSTFNISLNK-----EEMEFQPNQDEEDAH----- 386
Db 353 LPYTTAERFRENHPEPSTAKIEVPKNAAMPSTWSSVVIDEMADDASGHFSGAGNGKML 412
QY 387 -----AGIGRFLG-----LQSHDHPHPRANSRTKLLWPK 416
Db 413 DSSPLSVSVGTLNRVNTVASALKKFLSRDSDRPGSATPSQDPYKFPASASLSGAV 472
QY 417 RESLHGLPNKHAQKQNVGOENKA--WKLKAVD-----AFKSGGLY 459
Db 473 VGSATSAGKAPAGSLRITQCVIEEVDQATITSMRANPRPNVMDIFATSSGAGTSGPLQ 532
QY 460 QRPGYVSAQPQLSP-----TPMFFP-----LEPSAPS-----KLHSVTG-- 494
Db 533 PPAHSEPVDPISRPSPYNRAQSQVEPNLPPGGVDALLSTAPAGSGPLLSNATAPS 592
QY 495 --IDTKDKSLKTVSSGAKK---SPELLESB---GALMEHPVSQVRKTVFNLTDMPE 546
Db 593 SPVGSSEKSLYDPQKASRETVESMDLRSLTLLGDAAVQPEDEG-----DDFDK 642
QY 547 IPENHLKEPLOSPTNIHTTL 567
Db 643 LKAEREKEKLMFQKNLARTI 663

RESULT 5
QYVUM7
ID QYVUM7 PRELIMINARY; PRT; 535 AA.
AC QYVUM7;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-WAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG12327 protein.
GN CG12327.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Cherry J.M., Cavley S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisic A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Taveanli B.C., Pappu K., Mardon G.;
RT "deest" a Drosophila Homolog of Human Bestrophin.";
RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Faragaa V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003686; AAF54503.1; -
DR EMBL: AF218817; AAF32327.1; -
DR EMBL: AY061546; AAL29094.1; -
DR FlyBase; FBgn0040238; best.
DR InterPro; IPR000615; Wormfam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Wormfam_8; 1.
SQ SEQUENCE 721 AA; 79642 MW; B37CD96DE2F15B15 CRC64;

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.V., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hottin D., Houson K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson J.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.
"The genome sequence of *Drosophila melanogaster*."
Science 287:2185-2195 (2000).
DB EMBL: AE003531; AAP49648.1;
DR FlyBase; FBGN0036492; CG12327.
DR InterPro; IPR000615; Worm.fam.8.
DR Pfam; PF01062; DUF289; 1.
DR PRODOM; PD002803; Worm.fam 8; 1.
SQ SEQUENCE 535 AA; 62710 MW; 7C933A0E5068DF2F CRC64;

```

Query Match      28.4%; Score 886; DB 5; Length 535;
Best Local Similarity 40.8%; Pred. No. 4.7e-66;
Matches 177; Conservative 86; Mismatches 135; Indels 36; Gaps 8;

QY      1  MTITYSQVAN-ARLGSPSRLLLCWRGSIYKLLYGEFLICLLCYIIRFYIRALRTEQQ 59
Db      1  MTYSYTAEVATCSHGFCWKLLMEWRASIIYKIIVDLLAFPLSCFYFMAVIYRYALRVDK 60
QY      60  LMEKFLYCDSYIQLIPISVLGPGYVTVLVVTRWNOYENLPWDRMLSVSGPVEGKDE 119
Db      61  PVFEDIWYCHSYNLIPLSFVLGPGYVGIIRWNNQYIYVPDEDDPLAVYVSALVRQDE 120
QY      120  QSRLLRRTLRYANLGNVLIRLSYSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSL--- 176
Db      121  HGRLLMRRTIMRYVCLATVLSMISPIVKRRFPYDQLEIVGLLANEANIMKAMDVKPP 180
QY      177  PHNMFWPTWTFWPNANLSKMWALGGBIRDPDILLOSLLENEMWTLRTQCGHLYAYDWISIPLVY 236
Db      181  KHPKYMWPIVWAASIVTRABKEGRINWDDFSLKSMIDELNKFPRAGCNMLIHYDTISVPLVY 240
QY      237  TQVTVVAVYFSEFTCLTGROFLNPAKAYPGHELDLVVPVETFLQFPFVVGWGLKVAEOLIN 296
Db      241  TQVTVLAVISYFVASIFGHQWIDRDIKHYNIVSYFFLPFLSTLEFFPFMGWLKVAETLIC 300
QY      297  PFGEDDDDDFTNMTIVDRNLQVSLIADVEMHQDLPRMEPDMYWNKPEP-QSPYTAAS--AQ 353
Db      301  PFGDDDDDFELNMLIDRNLQVSYLIVDEMNDHEQLVRDQYWDVDEVPFAELPYAVESDRAE 360
QY      354  FRRASFP--MG-----STFNISLNKEEMEFQNOEDBEDAHAGIIGRFLGLQSHDHH 402

```

```

361 HPEASTARLGIPIKVPVMTKSEVSLNDFTEP-----DDEBYNPEVTIRF
403 PPRANSRTKLLWPK 416
408 -----ARREFSWSK 416

RESULT 6
O17206 PRELIMINARY; PRT; 612 AA.
AC O17206;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE C01B12.3 protein.
GN C01B12.3.
OS Caenorhabditis elegans.
OC Eukaryota, Metazoa, Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinatock L., Wilkinson-Sproat J., Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RS SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Scheet P., Maggi L.;
RA "The sequence of C. elegans cosmid C01B12.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RS SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025458; AAE70976.1; -
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 612 AA; 71031 MW; DFB43916541DD44 CRC64;

Query Match 26.1%; Score 814; DB 5; Length 612;
Best Local Similarity 33.6%; Pred. No. 6.7e-60;
Matches 193; Conservative

Qy 1 MTVTYSQVANARLGSFSSRLLLCWGRSGIYKLLYGFLIFLCYIIRFYRLALTEEQOL 60
Db 1 MTVTYSLDVASSSFFCYLLKLLFRWKGSIWKSVAELVWMLCLYAVLSYVYRCLLTMKQRA 60
Qy 61 MEFKLTYCDSYIQLIPISEVLGYPYVTVLVTRWNNQYENLPWPRLNSLVSGFVGKDEQ 120
Db 61 TFEELDCIFFDTYNSFNPIITFMWLGYYSAVFTRWQIIFDNIGWIDTPCLWITQYIKGETER 120
Qy 121 SLLRLRLTRYANLGNVLIRSYSTAYVKFPFSAQHLVQAGMTPTAEHKQLKLSLPHNM 180
Db 121 AKCVRENCIRYSILTQAVYVEDVAASVTRKFFPTFNHLVTAGLMTKEWAEFESIPSHAK 180
Qy 181 FWTVPWWFANLSKMWLGGRIKRPDILLQSLLENMNTLRTQCGHLVAYDWISIPLYVTQV 240
Db 181 YVQMPHFLWSMITLARDDEGMISDIIYVDLMKMRQFRVNTLSLTLDQWVPVPLVYTV 240

```


Db 63 KLIFKWSGLKAIYLDIVLWVCFYAFISVIVRYALDRSQDPTFRFMQFCNRRLLDYPI 122
 Qy 79 SFVLGFYVTVTRWNOYENLPWDLMSLVSGFVEGKDEOSRLRLRLTLRIYANLGNVL 138
 Db 123 NFMLGFFVTVTRWNOYENLPWDLMSLVSGFVEGKDEOSRLRLRLTLRIYANLGNVL 182
 Qy 139 ILRSVSTAVYKRFPSAHLVQAGFMTAPAHKOLEKLSLPHNNFWPWWFANLMSKAWLG 198
 Db 183 VFRDIHIGVRKRFPTLTWAAAGIMTSSELKYNVESRYAKYMLGFNFWNTFLNEARRE 242
 Qy 199 GRIRDPILLOSLNEMNTLRQCCHLYAYDMISLPLVTVVTVVAVYFFLTCVLNGROFL 258
 Db 243 GRIESAYTQNAIAEIRFRSGLSLWTVPIPLMYPQLVFMALHCYLVCLVSROFV 302
 Qy 259 NPAKAYPGHELDLVVPVFTFLQFPFYVGVKVAEOLINPFGBDDDDFTFNWIVDRNLQVS 318
 Db 303 INSDAVNTTEIDLVGVPFTIETIFVYMGWLVKVMDLNLPFFGEDEDDFCNFLIDRLNLTVA 362
 Qy 319 LLAVDENHODLPRMEPDWYNNKPEPPPYTAASAQFRASFMGSTFN----ISLNKEBME 374
 Db 363 MGIVDDSHDDGPFILEKDMFN--DTVSPLYSSAAQRNVNPFYGSATNADAQIPDDVRQIT 421
 Qy 375 FQPNQDEEDAHAGIIGFLGLQSHDHPHPRANSRTKLLPKRESLLHGLPKNHKAQK 434
 Db 422 MIPHLNEK-----LDQMYGKRT--NRPP--VQSVEL---KRDQRFSTG----- 459
 Qy 435 NVRGQEDNKWKLKAVDAFKSGPLQYRPGYVSAPOTPLSPTMFPPLPSAPSAPKLSHVTG 494
 Db 460 -----NNRKOTLEAKFNKLGAMQKQKHSKLTFTPEST-----KARASTD 501
 Qy 495 IDTKDKSLKTVSSGAKSKFELLSESDGALMEHPEVSQVRRKTVFNLTMDPPEIPENHLKE 554
 Db 502 IET---VSTPENTQKPCY-----SNPDCEIVEDEESQ-----DIPKVTPEPKQE 542
 Qy 555 P 555
 Db 543 P 543
 RESULT 9
 Q9VUM6 PRELIMINARY; PRT; 327 AA.
 ID Q9VUM6
 AC Q9VUM6;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE CG7259 protein.
 GN CG7259.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Maya A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Staden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 DR EMBL; AB03531; AAF9649.1; --
 DR FlyBase; Fegno036491; CG7259.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 SQ SEQUENCE 327 AA; 38594 MW; BE7CE3BC05370F46 CRC64;
 Query Match 23.8%; Score 742; DB 5; Length 327;
 Best Local Similarity 46.3%; Pred. No. 3.1e-54;
 Matches 142; Conservative 55; Mismatches 92; Indels 18; Gaps 4;
 Qy 68 YCDSYLIQIPISFVLGFGYVTVLVTWNOYENLPWDLMSLVSGFVEGKDEOSRLRLRT 127
 Db 3 YCEKNGSLIPSLFSLGFGFVRIVMNRWDQYTTIPWPDGIALIISTSIHSGSDRRVRRT 62
 Qy 128 LIRYANLGNLILRSVSTAVYKRFPSAHLVQAGFMTAPAHKOLEKLSLPHNNFWPWWFANLMSKAWLG 182
 Db 63 ILRYVCLQVIVFTWISPRVKERFPTTYQIIEAGFELLENEKKIIEAMDQAFPSYPKH--W 120
 Qy 183 VFWVWFANLMSKAWLGKIRDPILLOSLNEMNTLRQCCHLYAYDMISLPLVTVVTVV 242
 Db 121 MPVWAAISVMRRARENKIRDDYAVKTTIDELNQLRNGCGFLLYYDWISVPLVTVVTV 180
 Qy 243 AVYSFELCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFPFYVGVKVAEOLINPFGBDD 302
 Db 181 ATYSFELSVLGGQQWNSHDESDGTRIRRWFPILTVLQFPFYMGWLVKVAETLINPFGBDD 240
 Qy 303 DDFETNWTVDRLNQLVSLLAVIDEMHQDLPRMEPDWYNNK-----PEPQP-----PYTAAS 351
 Db 241 DDFELNWIIDRLNLTYSYCIYVDEMHEHPELVKDYWEVFPNEIPYVQPKMQNPAAST 300
 Qy 352 AQFRAS 358
 Db 301 AYMESQS 307
 RESULT 10
 O16779 PRELIMINARY; PRT; 1447 AA.
 ID O16779
 AC O16779; O16781;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Hypothetical 164.2 Kda protein.
 GN T21D12.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6339;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Woessner J.;
RC "The sequence of C. elegans cosmid T21D12.";
RT Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Waterston R.;
RC "Direct Submission.";
RT Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RL [3]
DR EMBL; AF016687; AAK72064.1; -.
DR HSSP; P56276; 1TLK.
DR InterPro; IPR003599; IG.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003600; IG_like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR Pfam; PF00047; IG; 3.
DR Pfam; PF00560; LRR; 12.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR PRODOM; PD002802; Worm_fam_8; 1.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 13.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 1447 AA; 164168 MW; 4BC1FE39E98E5B5 CRC64;

Query Match 23.0%; Score 717.5; DB 5; Length 1447;
Best Local Similarity 31.6%; Pred. No. 3.1e-51;
Matches 192; Conservative 97; Mismatches 228; Indels 91; Gaps 15;

Qy 1 MTITYTSQVANRLGSPSRLLLCWRSYKLYGFLIFLLCYIIRPTYR--LALTEHQ 58
Db 879 MTSVNSDVSSTSLTAPLRLQLRWRGSVMRELTFIFSCFAITSYRTNYFLSDEQ 938

Qy 59 QLMREKLTLYCDSYQTLPIPSVLGFTVTVTRWNNOVENPWPDLMSLVSGFVGKD 118
Db 939 RVFDWNFALFDQTLNYLPIPLFMGLFFVTIIVGRWDFLNLGWVDNTALLIATYIRGD 998

Qy 119 EQSRLRLRTLIARYNLGNVLILRSYSTAVYKFPSPAQHLVQAGFMTAPFHKLEKLSL-P 177
Db 999 EKSRIMRRTALRYLVLTQVLIFRDISLVYKVKFFLTETVVASGFMLESEKRYAALS LK 1058.

Qy 178 HNMFWVPWFANLSMKAWLGRIRDDPILLOSLNEMNTLRQCQHLVAYDWISIPLYVT 237
Db 1059 YPKYFLPQWCFSLYDARAQKIGADVMNLNLIKSVGDFRKLGLQLLNFWDVPLVYP 1118

Qy 238 QVTVVAVYSPFLTCLVGRQL-----NPAKAVPGHLDLVVP-VFTFLQFFVGVWL 288
Db 1119 QVFLAVRIYIFLCLIAKQSVLIDGKPKDNGSVTP-----FVFLMTSLQFVYVGM 1172

Qy 289 KVAEQLINPFGEDDDDDDFETNTWIVDRNLQVSLAVDEMHDQLRMBDFMYWNKPEFPPT 348

```

```

Db 1173 KVAESLMNPLGEDDDDFECNYLLDRNLAVGLAIVDS--SEAAPHYKDVFWGLAIADPLYS 1231
Qy 349 AASAOFRASFWGSTFNISLNKEEMEFQNOE--DEDAHAGIIG-----RFLGLQSH 399
Db 1232 SDTANIHNIPQIGSAATYETQENIIMRPHVDNTDMEDGDDVEGCNPRQLSRFVSVSV 1291
Qy 400 DHHPPRANSRTKLLWPKRESLLHGLPKNHKAAKNVRQOEONKAWLKXAVDAFKSGPLY 459
Db 1292 NRN---CESRTSL--SRNPITIIERLSRQFGSKYN-----1323
Qy 460 QRPGYVSAPQTPLSPTPMFFPLPSPAPSKLHSVTGIDTKSLKTVSSGAKKSFEILLSBS 519
Db 1324 -RPG-----KLFSSQFSINTAMGDN-----DIGSCASILGELAES 1359
Qy 520 D----GALMEHPVSVQVRKTVBFNLTDMPETPENHLKEPLEOSPNTIHHTTKHMDPYW 575
Db 1360 NKASQGLLTPEYAGSPRNTPIDMLTSVPE-----EDEBAQKTRTSVDRKWKEMVE 1412
Qy 576 ALENRDEA 583
Db 1413 NEKKREA 1420

RESULT 11
Q9VRW4
ID Q9VRW4 PRELIMINARY; PRT; 731 AA.
AC Q9VRW4;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE CG10173 protein.
GN CG10173.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., DeLor A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostek A., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jhalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Ringden K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sington K., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svendsen K., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

[illegible]

```

RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z92828; CAB07337.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 512 AA; 58380 MW; A49F8175CE1F0DCE CRC64;

Query Match
Best Local Similarity 19.8%; Score 616.5; DB 5; Length 512;
Matches 152; Conservative 94; Mismatches 217; Indels 53; Gaps 13;

QY 1 MTITVTSOVANARLGSFSLLCWRSIYKLYGFEFLIFLLCYIIRFYIRLALTEEQOL 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1 MTSVSNLDSVSSVFFNPKLLFRWRGWSKSIWSELVLLGYYLVWYIRVALITTEQKA 60

QY 61 MPEKLTLYCDSYI-OLIPISFVLGVFVTLVWTRWNNQYENLPWPDRLMSLVSGFVGKDE 119
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
61 GVRKYIDHLDQNDKCVLTFMLAFVTLIIVDRKNNMFANIGFIENATATLTKGTGEG 120

QY 120 OSRLRLTIRYANLGNVILRSVTAIVYKRPFSQHLVQAGFMTPAEHKQLEKLSL--- 176
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
121 DVLLAKRTIIRYLVLTQVLVFRDISLKVRRRFPNDHAIKAGFL--QDH---ESILGCD 175

QY 177 -PHNFWPWWFANLSKAWLGGRIIRDPILOSLLNEMNTLRTQCGHLYAYDWISIPLV 235
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
176 NGRNYMFWPNVSSAILQKLFEDGNIIPAFLNSVQEVKTPRSNNATLCNDYDWPPIA 235

QY 236 YQVVTVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVVFLOPFFVGVGLKVAEQLI 295
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
236 YQVVVFAVYVFFCLTRQHLDMEDT---KTIIDYYPILVTFQFFFMGMKVAEALL 292

QY 296 NPFGEDDDDFETNWIVDRLNQLVSLAVDEMHDLPMEPDMYNNKPEPPPY----- 347
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
293 NPLGEDDDDFECNYLIDRNATGMAIVNSKYSQVDEMLADKF-NDPSYAPYPERKVIDSG 351

QY 348 -----TAASAQFRASFMGSTFNISLNKEMEFQPNQOEDEDAHAGIIRFLGLQSH- 399
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
352 ADHALVGAQVTTLAEPNDIIDMMKVDLNSPIVVGRRNTNTSTIRRLSSAF-GRRSHS 410

QY 400 -DHHPPRANSTRKILMPKRESLLHGLPKNKAQKQNVGRGO-EDNKAWKLKAVDAFKSG- 456
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
411 VOHLGPEKPEPTS-----PFSQMAPQRYGAFELSNGNLNGLSGITSQS 456

QY 457 --PLYQRPGYISAP---QTPLSPPTMPFFLEPSAPS 487
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
457 HLPKLESESETDPLPAQOPVFLPRTLSEBPTSPT 492

RESULT 14
Q965X4 PRELIMINARY; PRT; 551 AA.
AC Q965X4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical protein Y37ELLAR.1.
GN Y37ELLAR.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;
RA Miller N., Maggi L.;
RT "The sequence of C. elegans cosmid Y37ELLAR.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024759; AAK68431.1; -.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 551 AA; 63476 MW; 8B65140751455F25 CRC64;

Query Match
Best Local Similarity 17.8%; Score 554; DB 5; Length 551;
Matches 146; Conservative 91; Mismatches 190; Indels 102; Gaps 13;

QY 1 MTITVTSOVANARLGSFSLLCWRSIYKLYGFEFLIFLLCYIIRFYIRLALTEEQOL 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
97 MTSVSNLDSVSSVTFNRLRLRWKSGIWSKSSSELMMWCFYFLIALLYHKVLPND-- 154

QY 61 MPEKLTLYCDSYIQLIPISFVLGVFVTLVWTRWNNQYENLPWPDRLMSLVSGFVGKDEQ 120
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
155 -----YDSAL-----TVNAIIIVSKCDE 172

QY 121 SLLRLTIRYANLGNVILRSVTAIVYKRPFSQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
173 ARLIRNRIYLVLAQILTPRDISIRVRRRFPNIDSIKKAGFLTNEBELLEGIDLAYNK 232

QY 181 FWPMWFWANLSKAWLGGRIIRDPILOSLLNEMNTLRTQCGHLYAYDWISIPLVTVQV 240
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
233 YWVFINWALSUNQANSKGYISAPGWSVLQIETKTPRGLATCNFQWCPIPIAYPOV 292

QY 241 TVAVYSFPLTCLVGRQFLNPAKAYPGHELD-----LVVPVFTFLOPFFVGVGLKVAEQLIN 296
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
293 FFAVRIYFLCLITRQYIR----VFNKELEPVQMFIRPFIIFICIVGMKVAEALIN 348

QY 297 PFGEDDDDFETNWIVDRLNQLVSLAVDEMHDLPMEPDMYNNKPEPPPYTAASQA- 355
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
349 PLGEDDDDFESNFLIDKNIFTGMRIYVQFDEAPLEFDT--FSDPGAMPIYSEESQKNYQ 406

QY 356 RASFMGSTFNISLNKEMEFQPNQOEDEDAHAGIIRF-LG-LOSHDHHPPRANSTRKIL 413
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
407 NGALVGSVSNVTL-----AQFDENITWVPVAPRLSIGDINAH-----RSSIRRRFK 453

QY 414 WPKRESLLHGLPKNKAQKQNVGRGOEDNKAWKL-----KAVDAFKSGPLYQRPGYISA 467
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
454 SGSSQNCSTRANSFREPKNVLRQPSDGSNDAYELDYETVKSVDLVLAS----- 502

QY 468 PQPLSPPTMPFFLEPSAPSKLH-SVTGIDTKQSKLTKTVSSGAKKSFEL 515
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
503 -----SPTKPGKFFTSKTVDENDIDIPVSLIFAKKSLGL 537

RESULT 15
O17674 PRELIMINARY; PRT; 400 AA.
ID O17674;
AC O17674; O46026;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE C49A1.2 protein.
DE C49A1.2 protein.
GN C49A1.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]_TaxID=6239;
RP SEQUENCE FROM N.A.
RA Matthews L.;

```

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirschen J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin C., Roodpr A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Kershaw J.,
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 283221; CAB05710.1; -
DR EMBL; 282095; CAB05710.1; JOINED.
DR EMBL; 282095; CAB05029.1; -
DR EMBL; 283221; CAB05029.1; JOINED.
DR InterPro; IPR000615; Worm_fam_9.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
SQ SEQUENCE 400 AA; 46827 MW; D1CDF601E25E2D01 CRC64;

Query Match 15.9%; Score 495; DB 5; Length 400;
Best Local Similarity 32.9%; Pred. No. 2.7e-33;
Matches 119; Conservative 63; Mismatches 130; Indels 50; Gaps 8;

QY 1 MTITVTSOVANARLGSFGRLLLCWGSIVKLYGFLIFLCVYIIRFYRLAL-----T 55
DB 1 MTITYRAVSTESINFLKIIIRHWGSLAKSVQEFWLFYVYVCAVYRCVLFIGWD 60

QY 56 BEQQLMFELTYLDCSYIQLIPISFVLGPFYVTVTRMNOYENLPWDRMLSLVSGFVE 115
DB 61 DQPKFVEILSTHDMY---IPLEFMLGPFVTVIDRWKSPQNIPIYESCAPAVSAALP 117

QY 116 GKD-----EQSRLRLRIYRANLGNVLILRSVSTAVYKRPSPAGHLVQAGFMTPA 166
DB 118 GHGALEVSAGMDTKLTARTIIRYLVLSQILLPREFSVKVKRFRVDRSLVDSKFLTEN 177

QY 167 EHKOL--EKLSPHNMFWVPWVPANL-----SMKAWLGGIRDPILLQSLNE 213
DB 178 ELTELSEELTKRYDSYILPINWAFSILRTEKLNPNQPMNDAW-----NV 222

QY 214 MNTLRTQCGHLYAYDMISIPLYVTQVTVVAVYSFPLTCLVGRQFLNPAKAYPGHDLVV 273
DB 223 INDQVKLTLLRNGDFIPIPLAYPQAVFLAVRFVLCVLFTRQHLDS---DHAIDYFF 279

QY 274 PVFIFLQFFVYVGLKVAEQLINPGEDEDDDEETWIVDRNLQVSLAVDENHQDLPRME 333
DB 280 PLLTSFQFIFIVGWMKVAEILNPNMGEDDDDFELNYVIDKNFYIGMTIVDS--KDIELTE 337

QY 334 PD 335
DB 338 ND 339

Search completed: July 10, 2003, 12:26:17
Job time : 36.0242 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 26.8587 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-5
Perfect score: 2350
Sequence: 1 MTITYSQVANRLGFSRL.....QLLPSSVEPLWAPPSTSA 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2350	100.0	435	20	AA19954
2	1535	65.3	585	20	AA19953
3	1297	55.2	551	20	AA19955
4	763.5	32.5	721	22	ABG2213
5	685	29.1	535	22	ABG3121
6	543.5	23.1	327	22	ABG3135
7	493	21.0	731	22	ABG4914
8	450	19.1	368	22	ABG08609
9	361	15.4	314	22	AA190085
10	338	14.4	701	22	ABG08608

11	239	10.2	251	22	AA170067	Human secreted pro
12	239	10.2	251	23	ABG65506	Human albumin fusi
13	238	10.1	99	22	AAU22652	Novel human colon
14	238	10.1	99	22	AAU22658	Human digestive sy
15	204.5	8.7	377	22	ABG6411	Human human diago
16	174	7.4	30	22	AA19956	Human CGICE partia
17	151	6.4	118	22	ABG04059	Human musculooske
18	144	6.1	108	22	AAU90422	Human immune/haema
19	141.5	6.0	72	22	AAO13117	Human polypeptide
20	138	5.9	76	22	AAO13117	Human ORFX ORF3036
21	137	5.8	60	22	ABG3675	Human musculooske
22	136.5	5.8	78	22	AAU86740	Human immune/haema
23	136.5	5.8	102	22	AAU95008	Human protein sequ
24	135	5.7	71	23	ABG79293	Human prostate spe
25	135	5.7	93	22	AAO13000	Human polypeptide
26	134	5.7	43	22	AAU85912	Human immune/haema
27	134	5.7	65	22	ABG75572	Human colon cancer
28	133	5.7	278	23	ABP51443	Human MDDR SEQ ID
29	132	5.6	76	21	AAU45219	Human secreted pro
30	131.5	5.6	90	22	AAU30514	Novel human secret
31	130.5	5.6	63	21	AAU63223	Gene 43 human secr
32	130.5	5.6	68	21	AAU44753	Gene 23 human secr
33	130.5	5.6	71	21	AAU44759	Gene 50 human secr
34	130.5	5.6	72	21	AAU45218	Human secreted pro
35	130.5	5.6	74	21	AAU51464	Human secreted pro
36	130.5	5.6	80	22	AAO12799	Human polypeptide
37	130.5	5.6	40	22	AAU75544	Human colon cancer
38	130	5.5	76	22	AAU69856	Human prostate CDN
39	130	5.5	76	22	AAU69933	Human prostate CDN
40	130	5.5	76	22	AAU61211	Human prostate-spe
41	130	5.5	76	22	AAU61211	Human prostate-spe
42	130	5.5	76	22	AAU61211	Human prostate-spe
43	130	5.5	76	22	AAU61211	Human prostate-spe
44	130	5.5	76	22	AAU61211	Human prostate-spe
45	130	5.5	107	22	AAO12697	Human polypeptide

ALIGNMENTS

RESULT 1
AA19954
ID AA19954 standard; Protein: 435 AA.
AC AA19954;
DT 22-NOV-1999 (first entry)
XX Human CGICE short form protein sequence.
DE CGICE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX Homo sapiens.
XX WO9943695-A1.
XX 02-SEP-1999.
XX 22-FEB-1999; 99WO-US03790.
XX 25-FEB-1998; 98US-0075941.
XX 18-DEC-1998; 98US-0112926.
XX (MERI) MERCK & CO INC.
XX (UYUP-) UNIV UPPSALA.
XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX WPI; 1999-540560/45.
XX N-FSDB; AA221228.
XX Human and mouse polynucleotides encoding CGICE polypeptides

XX PS Claim 7; Fig 5; 67pp; English.

XX CC The present sequence represents the human CGICE protein. When the CGICE

XX CC gene is mutated it is responsible for Best's macular dystrophy (BMD).

XX CC Polynucleotides encoding CGICE are useful for diagnosing whether a

XX CC patient carries a mutation in the CGICE gene. Normal and mutated

XX CC CGICE proteins are useful for identifying activators and/or inhibitors

XX CC of these proteins, in order to treat BMD. The CGICE gene offers a

XX CC simpler and cheaper method of diagnosing BMD without the need for the

XX CC presence of the patient. The gene may also be useful to discovering

XX CC the genetic cause of age-related macular dystrophy.

XX SQ Sequence 435 AA;

Query Match 100.0%; Score 2350; DB 20; Length 435;

Best Local Similarity 100.0%; Pred. No. 3.3e-237;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

DB 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

QY 61 MPEKLTLYCDSYIQLIPISFVLGFFVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

DB 61 MPEKLTLYCDSYIQLIPISFVLGFFVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

QY 121 GRLLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

DB 121 GRLLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240

DB 181 FWVPWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240

QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLQFFFYVGLKVSRLALLGWR 300

DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLQFFFYVGLKVSRLALLGWR 300

QY 301 HGQRHGQQLLETRMOCQERKVSRYESSQAWRTVPVTPATREAEAGESLEPGRRRLWQS 360

DB 301 HGQRHGQQLLETRMOCQERKVSRYESSQAWRTVPVTPATREAEAGESLEPGRRRLWQS 360

QY 361 SSSSTPLERMMLIRPTGLSTGICRCPCWLMWRCRTCLGWSRTCTGISPSHSPTQLPP 420

DB 361 SSSSTPLERMMLIRPTGLSTGICRCPCWLMWRCRTCLGWSRTCTGISPSHSPTQLPP 420

QY 421 SSVPEPLWAPPSTSA 435

DB 421 SSVPEPLWAPPSTSA 435

RESULT 2

AAV29953

ID AAV29953 standard; Protein; 585 AA.

XX AC AAV29953;

XX DT 22-NOV-1999 (first entry)

XX DE Human CGICE long form protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 121

XX FT /note= "encoded by GGC"

XX PN WO9943695-A1.

XX XX

PD 02-SEP-1999.

XX 22-FEB-1999; 99WO-US03790.

XX 25-FEB-1998; 98US-0075941.

PR 18-DEC-1998; 98US-0112926.

XX (MERI) MERCK & CO INC.

PA (UYUP-) UNIV UPPSALA.

XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;

XX WPI; 1999-540560/45.

DR N-PSDB; AA221227.

XX Human and mouse polynucleotides encoding CGICE polypeptides -

XX Claim 7; Fig 3; 67pp; English.

XX The present sequence represents the human CGICE protein. When the CGICE

XX gene is mutated it is responsible for Best's macular dystrophy (BMD).

XX Polynucleotides encoding CGICE are useful for diagnosing whether a

XX patient carries a mutation in the CGICE gene. Normal and mutated

XX CGICE proteins are useful for identifying activators and/or inhibitors

XX of these proteins, in order to treat BMD. The CGICE gene offers a

XX simpler and cheaper method of diagnosing BMD without the need for the

XX presence of the patient. The gene may also be useful to discovering

XX the genetic cause of age-related macular dystrophy.

XX SQ Sequence 585 AA;

Query Match 65.3%; Score 1535; DB 20; Length 585;

Best Local Similarity 99.7%; Pred. No. 1.2e-151;

Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

DB 1 MTITVTSQVANARLGSFRLLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQQL 60

QY 61 MPEKLTLYCDSYIQLIPISFVLGFFVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

DB 61 MPEKLTLYCDSYIQLIPISFVLGFFVTLVVTWNNQYENLPWDRMLSLVSGFVEGKDRQ 120

QY 121 GRLLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

DB 121 GRLLRRTLIIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

QY 181 FWVPWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240

DB 181 FWVPWVFWANLGMKAWLGGRIIDPILLOSLLNEMNTLRQCGLHYAYDWISIPLVYTVQV 240

QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLQFFFYVGLKVSRLALLGWR 290

DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHLDLVVPVFTFLQFFFYVGLKVSRLALLGWR 290

RESULT 3

AAV29955

ID AAV29955 standard; Protein; 551 AA.

XX AC AAV29955;

XX DT 22-NOV-1999 (first entry)

XX DE Mouse CGICE protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Mus sp.

XX PN WO9943695-A1.

XX XX

```

PD 02-SEP-1999.
XX
XX 22-FEB-1999; 99WO-US03790.
XX
XX 25-FEB-1998; 98US-0075941.
XX 18-DEC-1998; 98US-0112926.
XX
XX (MERI ) MERCK & CO INC.
XX (UYUP-) UNIV UPPSALA.
XX
XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX
XX WPI; 1999-540560/45.
XX N-PSDB; AAZ21229.
XX
XX Human and mouse polynucleotides encoding CGICE polypeptides
XX
XX Claim 7; Fig 8; 67pp; English.
XX
XX The present sequence represents the mouse CGICE protein. When the CGICE
XX gene is mutated it is responsible for Best's macular dystrophy (BMD).
XX Polynucleotides encoding CGICE are useful for diagnosing whether a
XX patient carries a mutation in the CGICE gene. Normal and mutated
XX CGICE proteins are useful for identifying activators and/or inhibitors
XX of these proteins, in order to treat BMD. The CGICE gene offers a
XX simpler and cheaper method of diagnosing BMD without the need for the
XX presence of the patient. The gene may also be useful to discovering
XX the genetic cause of age-related macular dystrophy.
XX
XX Sequence 551 AA;
SQ
Query Match 55.2%; Score 1297; DB 20; Length 551;
Best Local Similarity 82.4%; Pred. No. 9.3e-127;
Matches 239; Conservative 23; Mismatches 28; Indels 0; Gaps 0;
Qy 1 MTITYTSOVANARLGSGFRLLLCWRGSIYKLYGFEFLFLCYIIRFYRLALTEBOOL 60
Db 1 MTITYTSOVANARLGSGFRLLLCWRGSIYKLYGFEFLFLCYIIRFYRLALTEBOOL 60
Qy 61 MFPEKLTLYCDSYIOLIPISFVLGFGYVTLVTRWNNQYENLWPDRLMSLVSGFVGKDBE 120
Db 61 LFEKALYCDYSYIOLIPISFVLGFGYVTLVTRWNNQYENLWPDRLMTQVSSFVGKDBE 120
Qy 121 GRLLRRTLRVANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Db 121 GRLLRRTLRVANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Qy 181 FWVPWVWPFANLSMKAWLGGRIRDPILLOSLLNEMNTLRTQCCHLYAYDWISIPLVYTVV 240
Db 181 FWVPWVWPFANLSMKAWLGGRIRDPILLOSLLNEMNTLRTQCCHLYAYDWISIPLVYTVV 240
Qy 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYVGLVKV 290
Db 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYVGLVKV 290
RESULT 4
ABB62213
ID ABB62213 standard; Protein; 721 AA.
XX
XX ABB62213;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 13431.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX

```

```

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06316.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 13431; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 721 AA;
SQ
Query Match 32.5%; Score 763.5; DB 22; Length 721;
Best Local Similarity 49.4%; Pred. No. 1e-70; 80; Indels 23; Gaps 5;
Matches 152; Conservative 53; Mismatches 53;
Qy 1 MTITYTSOVANAR-LGSGFRLLLCWRGSIYKLYGFEFLFLCYIIRFYRLALTEBOQ 59
Db 1 MTITYTSOVANAR-LGSGFRLLLCWRGSIYKLYGFEFLFLCYIIRFYRLALTEBOQ 59
Qy 60 LMFEKLTLYCDSYIOLIPISFVLGFGYVTLVTRWNNQYENLWPDRLMSLVSGFVGKDBE 119
Db 61 ETFEAVQYCDYSYIOLIPISFVLGFGYVTLVTRWNNQYENLWPDRLMSLVSGFVGKDBE 120
Qy 120 QGRLRRTLRVANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 177
Db 121 RGRMRRTLRVANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Qy 178 H-NMFVWVWPFANLSMKAWLGGRIRDPILLOSLLNEMNTLRTQCCHLYAYDWISIPLVY 236
Db 181 RPSKHLVIVWNAASITTRARKEGRIRDDPAVKTIIDELNKFQGCGLLSYDTISVPLVY 240
Qy 237 TQVTVVAVYSFELTCLVGRQ-----FLNPAKAYPGHELDLVVPVFTLQFFYV 285
Db 241 TQVTVVAVYSFELTCLVGRQ-----KVDLYFVFTLQFFYV 292
RESULT 5
ABB63121
ID ABB63121 standard; Protein; 535 AA.
XX
XX ABB63121;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 16155.
XX
XX

```

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07224.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 16155; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 535 AA;
 SQ
 Query Match 29.1%; Score 685; DB 22; Length 535;
 Best Local Similarity 42.9%; Pred. No. 1.1e-62;
 Matches 126; Conservative 66; Mismatches 98; Indels 4; Gaps 2;
 QY 1 MTITYTSQVAN-ARLGSFRLLLCMRGSYKLLYGFLFLCYIIRIYRIALTEEQ 59
 DB 1 MTVSYTAEVATCSHFQCFWKLWRWFASIKIIVDLLAFSLCFYFMAVIYVALRDVX 60
 QY 60 LMFELKLYCDYSYIOLIPISFVLGFFVTLVTRWNNQYENLWPDRLMSLVSGFVEGKDE 119
 DB 61 PVFEDIVMYCHSYNLIPLSFVLGFFVYGIIRWNNQYITVPWDPPLAVVYSALVRGQDE 120
 QY 120 QGRLLRRTLIRYANLGNVLILRSVSTAVYKRPFSQHLVQAGFMTPAEHKQLEKLSL--- 176
 DB 121 HGRLLRRTIMRVCLATLWLSMSVPIKRRPFTDQLLEVLGNANINMKAMDVKFP 180
 QY 177 PHNMFVWPVWPVWPFANLSMKAWLGGRIIRDPIILLOSLLNEMNTLTQCCHLYAYDWISIPLVY 236
 DB 181 KHPKYWMPVWAASIVTRARKEGRWDWDFSLKSMIDELNKRAGCNMLHYDTISVPLVY 240
 QY 237 TOVTVVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVVFVFTPLQFFFFYVGMWKV 290
 DB 241 TOVTVLAVYSYFVASIFGHQWIDRIDIKHYNINVSYYFFLFTLEFFFFGMWKV 294
 RESULT 6
 ABB63135
 ID ABB63135 standard; Protein; 327 AA.
 XX
 AC ABB63135;
 XX
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 16197.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07238.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 16197; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 327 AA;
 SQ
 Query Match 23.1%; Score 543.5; DB 22; Length 327;
 Best Local Similarity 44.3%; Pred. No. 3.9e-48;
 Matches 101; Conservative 48; Mismatches 72; Indels 7; Gaps 2;
 QY 68 YCDSYIOLIPISFVLGFFVTLVTRWNNQYENLWPDRLMSLVSGFVEGKDEQGRLLRRT 127
 DB 3 YCKNGSLIPLSFVLGFFVTRVNRWWDQYTTIPWPGIALISTSIHSGDDRAVMRT 62
 QY 128 LIRYANLGNVLILRSVSTAVYKRPFSQHLVQAGFMTPAEHKQLEKLSL-----SLPHNMF 182
 DB 63 ILRVYCLQCVITFTWISPRVKRPFYTTQITTEAGFLENEKKIIEAMDQAFPSYPKH--W 120
 QY 183 VPWFWFANLSMKAWLGGRIIRDPIILLOSLLNEMNTLTQCCHLYAYDWISIPLVYTVVTV 242
 DB 121 MPITWASIVWRARENKIRDDYAVKTIIDELNQRNGCGLLYDMSIPLVYTVVTV 180
 QY 243 AVYFFFTCLVGRQFLNPAKAYPGHELDLVVVFVFTPLQFFFFYVGMWKV 290
 DB 181 ATYSFFFLPSVLGQQWNSHSDQTRRRRWFPIITLVLFQFFYVGMWKV 228
 RESULT 7
 ABB64914
 ID ABB64914 standard; Protein; 731 AA.
 XX
 AC ABB64914;
 XX
 DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 21534.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 XX WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PF 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-PSDB; ABL09017.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT Disclosure; SEQ ID NO 21534; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL161175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 731 AA;
 SQ Query Match 21.0%; Score 493; DB 22; Length 731;
 Best Local Similarity 45.7%; Pred. No. 2.3e-42;
 Matches 101; Conservative 36; Mismatches 76; Indels 8; Gaps 3;
 QY 78 ISFVLGFTVTVTRWQYENLPWPDRMLSLVSGFVEG-----KDEQRLRLRLIRYA 132
 Db 1 MSFVLGFYVNLVVRWWEQYELLPPDPTLALFISAAIENSNGVNNETGRLMRRNINRYM 60
 QY 133 NLGNVLILRSVSTAVYKRPFAQHLVQAGFTPTAEHKOLEKLS--LPHNMFVWPWMPAN 190
 Db 61 VLAVITLQRLSLVRKRPFTQHLVDAGLHSEMKIFEALNQKSPSKYMPFLVWATN 120
 QY 191 LSMKANTGGRIRDPILQSLNEMNTLTQCCHLYAYDWISIPLYTQVTVVAVYSFELT 250
 Db 121 IINRARKDGLIASDHIVQTLVELSDIRRLGLLIGYDTCVPLVYTVQVTVLVLYTYIA 180
 QY 251 CLVGRQFLNPAKAPGHE-LDLVVPVFTFLOFFVYVGLKV 290
 Db 181 ALLGRQMLPNVLDNRSGREDPDLPFLPTVLOFVYVGLKV 221
 RESULT 8
 ABG08609
 ID ABG08609 standard; Protein; 368 AA.
 XX ABG08609;
 XX 13-FEB-2002 (first entry)
 DT Novel human diagnostic protein #8600.
 DE

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 XX WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PF 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX N-PSDB; AAS72796.
 DR New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT Claim 20; SEQ ID NO 38968; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 368 AA;
 SQ Query Match 19.1%; Score 450; DB 22; Length 368;
 Best Local Similarity 44.6%; Pred. No. 2.8e-38;
 Matches 91; Conservative 27; Mismatches 66; Indels 20; Gaps 4;
 QY 162 FMTPEAKHOLEKLSLPHNMFVWPWMPANLSMKALGGRIRDPILQSLNEMNTLTQC 221
 Db 1 FMTTDERKLFNHLKSPHLKYWVPFIWFGNLATKARNEGRINDSDLOSMTMERNYRSC 60
 QY 222 GHLYADYDWISIPLYTQVTVVAVYSFELTCLVGRQFLNPAKAPGHELDLVVPVFTFLOF 281
 Db 61 SLFLGYDWVGIPLVYTVQVTVLAVYVTFACILGQFLDPTKGYAGHDLDFIPIFTLQF 120
 QY 282 FFYVGLKVLGSLRALLGWRHQRGHGQOLLETRMQCOEKVKSRVSSQAWWRTVPVIPATR 341
 Db 121 FFYAGWLKV--AEQLI---NPFGEDDDDFTNW-CIDRLNQ-----VSLLA 160
 QY 342 EAEAGESLEPGRRLMWQSSSTP 365
 Db 161 VDEHMSLPKMKKDIYWDDSAARP 184

RESULT 9
 AAB70085
 ID AAB70085 standard; Protein; 314 AA.
 XX
 AC AAB70085;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Human secreted protein-related polypeptide #6.
 XX
 KW Human; secreted protein; immunomodulatory; anti-sclerotic;
 KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
 KW antialzheimer's; antiparkinsonian; antimicrobial; vulnery; gene therapy;
 KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
 KW neurological; infection.
 XX
 OS Homo sapiens.
 XX
 PN WO200112776-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 15-AUG-2000; 2000WO-US22350.
 XX
 PR 16-AUG-1999; 99US-0148759.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
 XX
 DR WPI; 2001-244245/25.
 XX
 PT Nucleic acids encoding 18 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Page 377-378; 380pp; English.
 XX
 CC The present sequence is given in a specification relating to nucleic
 CC acid molecules encoding 18 novel human secreted proteins. The nucleic
 CC acids and proteins may be used in the prevention, diagnosis and
 CC treatment of diseases including immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. The
 CC nucleic acid molecules may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples. The polypeptides may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators of
 CC protein expression and activity.
 XX
 SQ Sequence 314 AA;
 Query Match 15.4%; Score 361; DB 22; Length 314;
 Best Local Similarity 43.6%; Pred. No. 4.8e-29;
 Matches 88; Conservative 14; Mismatches 78; Indels 22; Gaps 4;
 QY 172 EKLSLPHNMFVWVWFWANLSKMWLGGRIRPILLOSLNENWTLRTGCHLYAYDWIS 231
 DB 13 ESLSKDFNKYVPCVWFNTLAAQARDGRDRIIDIALCLLEELNKYRKACMLFHYDWIS 72
 QY 232 IPLVTVQVTVAVYGFLLTCLVGRQFLNP- - - - AKAY- - - - PGHE- - - - LDLVVPVF 276
 DB 73 IPLVTVQVTVAVYGFLLTCLVGRQFLNP- - - - AKAY- - - - PGHE- - - - LDLVVPVF 276
 QY 277 TPLQFFFYVGMKVLGSLRALLGWRHQRGCHGOOLLETRMQCQKYSRV- - - - ESSQ 329
 DB 133 TLLQFFFYVGMKVLGSLRALLGWRHQRGCHGOOLLETRMQCQKYSRV- - - - ESSQ 329
 QY 330 AWRRTFVTPATREAEAGSLEP 351
 DB 193 YWDEDOQPQPPYTVATAESLRP 214
 RESULT 10
 AAB70085
 ID AAB70085 standard; Protein; 701 AA.
 XX
 AC AAB70085;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #8599.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS72795.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 38967; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAB70085-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 701 AA;
 Query Match 14.4%; Score 339; DB 22; Length 701;
 Best Local Similarity 26.7%; Pred. No. 3.7e-26;

RESULT 13	
AAU22652	
ID	AAU22652 standard; Protein; 99 AA.
XX	
AC	AAU22652;
XX	
DT	17-DEC-2001 (first entry)
XX	
DE	Novel human colon associated polypeptide #185.
XX	
KW	Human; colon cancer; congenital abnormality; infection; colitis;
KW	inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW	intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW	sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW	cytostatic.
XX	
OS	Homo sapiens.
XX	
PN	WO20015302-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01240.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.

sequences encoding for them. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of disorders of the colon including colon cancer, congenital abnormalities (e.g. atresia and stenosis), bacterial and viral infections, inflammatory bowel disease (IBD), neoplastic cell disorders (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis, colonic inflammation, diarrhoea and dysentery, malabsorption syndromes (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases. The polynucleotide sequences of the invention can also be used in gene therapy. AAU22468-AAU22701 represent the novel human colon associated polypeptides of the invention.

Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 99 AA;

Query Match 10.1%; Score 238; DB 22; Length 99;
Best Local Similarity 55.4%; Pred. No. 7.5e-17;
Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0

Qy 1 MTITYTSOVANARLGSFRLLCWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQOL 60
Db 13 MTIVSYTLKVAERFGFGSLLRWRGSIYKLYGFLIFLLCYIIRFYRLALTEEQRY 72
Qy 61 MFPEKLTLYCDSYIQLIPISFVLG 83
Db 73 VYAQVARYCNRSDLIPLSFVLG 95

RESULT 14
AAM92658
ID AAM92658 standard; Protein; 99 AA.
XX AC AAM92658;
XX DT 05-NOV-2001 (first entry)
XX DE Human digestive system antigen SEQ ID NO: 2007.
XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX KW digestive system disorder; Meckel's diverticulum.
XX OS Homo sapiens.
XX PN W0200155314-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01324.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.

20-OCT-2000; 2000US-0240950.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251889.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465567/50.
N-PSDB; AAS39532.
Isolated polypeptide for treating, preventing and/or prognosing disorders related to the colon including colon cancers and also for testing and detection e.g. diagnosis -
Claim 11; SEQ ID No 429; 562pp; English.
The present invention relates to the isolation of novel human colon associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic

QY 61 MPEKLTLYCDSYIOLIPISFVLG 83
 Db 73 VYAQVARYCNSADLIPLSFVLG 95

RESULT 15
 ABG06411
 ID ABG06411 standard; Protein; 377 AA.

AC ABG06411;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #6402.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS70598.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations,
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20; SEQ ID No 36770; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 377 AA;

Query Match 8.7%; Score 204.5; DB 22; Length 377;
 Best Local Similarity 30.1%; Pred. No. 1.5e-12;
 Matches 65; Conservative 26; Mismatches 64; Indels 61; Gaps 8;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLLYGEFLIFLCYIIIRPIYR----LALTE 56

Db 51 MTVTYARVANARFGFSQQLLLWEGSIYKLLWRLLCFLGFYMAISAAAYRAGLQICAAE 110
 QY 57 EQOLMFEKLTLYCDSYIOLIPISFVLGFIYTLVTRWNNQYENLPWPDRLMSLVSGFVEG 116
 Db 111 GKRYFEKLTLYCDH----MPASSLS-----PSCILLMLSEEKQS 146
 QY 117 KDEGRLRLRTLLRYANLG-----NVLLRSYSTAVYKRPSPSAQHLVQAGF---MTPA 166
 Db 147 LAQENAGLRERWGRPEGETPGLTAKKLLILQS-----QLEQLQENFRALVRPR 196
 QY 167 EHKOLEKLSLPH--NMFWVP-----WVWFA 189
 Db 197 FOALILQAEIILHVRKFWTPNNSGSGPEAKVIYDINFA 232

Search completed: July 10, 2003, 12:24:12
 Job time : 28.8587 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:22:34 ; Search time 9.96817 Seconds
(without alignments)
1283.984 Million cell updates/sec

Title: US-09-622-964-5

Perfect score: 2350

Sequence: 1 MTITYTSQVANARLGSFRL.....QLPPSSVEPPLWAPPSTSA 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
- 1: /cgm2_6/ptodata/1/1aa/5A.COMB.pep.*
 - 2: /cgm2_6/ptodata/1/1aa/5B.COMB.pep.*
 - 3: /cgm2_6/ptodata/1/1aa/6A.COMB.pep.*
 - 4: /cgm2_6/ptodata/1/1aa/6B.COMB.pep.*
 - 5: /cgm2_6/ptodata/1/1aa/PTUS.COMB.pep.*
 - 6: /cgm2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	5.5	76	4	US-09-605-785-575
2	111	4.7	52	4	US-09-227-357-537
3	97	4.1	681	2	US-08-272-255-6
4	97	4.1	681	2	US-08-272-255-9
5	97	4.1	681	5	PCT-US95-08565-6
6	97	4.1	681	5	PCT-US95-08565-9
7	95.5	4.1	582	3	US-08-906-865-3
8	95.5	4.1	582	4	US-09-129-668-3
9	95	4.0	342	3	US-08-785-928-1
10	95	4.0	342	3	US-08-728-603-17
11	94	4.0	681	3	US-08-964-268-5
12	92.5	3.9	353	1	US-08-229-418-2
13	92.5	3.9	353	2	US-08-932-761A-2
14	92.5	3.9	353	4	US-09-307-912-2
15	92.5	3.9	353	5	PCT-US95-04464-2
16	91.5	3.9	1692	4	US-09-263-933-4
17	91.5	3.9	2307	4	US-09-263-933-2
18	89.5	3.8	1692	4	US-09-263-933-18
19	89.5	3.8	2013	1	US-08-324-977-12
20	89.5	3.8	2013	2	US-08-384-616-12
21	89.5	3.8	2013	2	US-08-904-686A-12
22	89.5	3.8	2013	4	US-09-315-850-12
23	89.5	3.8	2307	4	US-09-263-933-16
24	89.5	3.8	2620	1	US-08-324-977-32
25	89.5	3.8	2620	2	US-08-384-616-32
26	89.5	3.8	2620	2	US-08-904-686A-32
27	89.5	3.8	2620	4	US-09-315-850-32

28	89.5	3.8	2621	1	US-08-324-977-36	Sequence 36, Appl
29	89.5	3.8	2621	2	US-08-384-616-36	Sequence 36, Appl
30	89.5	3.8	2621	2	US-08-904-686A-36	Sequence 36, Appl
31	89.5	3.8	2621	4	US-09-315-850-36	Sequence 2, Appl
32	89.5	3.8	3010	1	US-08-324-977-2	Sequence 2, Appl
33	89.5	3.8	3010	1	US-08-324-977-14	Sequence 14, Appl
34	89.5	3.8	3010	2	US-08-384-616-2	Sequence 2, Appl
35	89.5	3.8	3010	2	US-08-384-616-14	Sequence 14, Appl
36	89.5	3.8	3010	2	US-08-904-686A-2	Sequence 2, Appl
37	89.5	3.8	3010	2	US-08-904-686A-14	Sequence 14, Appl
38	89.5	3.8	3010	4	US-09-315-850-2	Sequence 2, Appl
39	89.5	3.8	3010	4	US-09-315-850-14	Sequence 14, Appl
40	88.5	3.8	1692	4	US-09-263-933-11	Sequence 11, Appl
41	88.5	3.8	2307	4	US-09-263-933-9	Sequence 9, Appl
42	86.5	3.7	3010	4	US-09-014-416-3	Sequence 3, Appl
43	86	3.7	509	3	US-08-845-546-2	Sequence 2, Appl
44	85	3.6	294	1	US-08-142-439A-4	Sequence 4, Appl
45	85	3.6	294	2	US-08-869-477-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-605-785-575
; Sequence 575, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 575
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-575

Query Match 5.5%; Score 130; DB 4; Length 76;
Best Local Similarity 63.2%; Pred. No. 4.1e-06;
Matches 24; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 318 QERKVSRYSSQAWRTTPVTPATREAEAGESLEPGRRR 355
Db 3 KSRFTKTKITQAWRAPVPTREAEAGESLEPGRLR 40

RESULT 2

US-09-227-357-537
; Sequence 537, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.

```

; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-537

Query Match 4.7%; Score 111; DB 4; Length 52;
Best Local Similarity 72.4%; Pred. No. 0.00022;
Matches 21; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 328 SQAWRTTPVIPATREAGESLEPGRRRL 356
Db | | | | | | | | | | | | | | | |
   23 SWLWHTPPVVPATWEAEVGSLEPGRLRL 51

RESULT 3
US-08-272-255-6
; Sequence 6, Application US/08272255
; Patent No. 5824859
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,255
; FILING DATE: 08-JUL-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-272-255-6

Query Match 4.1%; Score 97; DB 2; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

Qy 123 LLRRRLIRYANLGNV-----LILRSYSTAVYKEFPQAQLVQAGFMTPAEHKOLEKLS 175
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   267 LVRIQVAVANEAGNAGESVNLFKLSGLREYSRIISFNH-----PSSH-----ERPL 315

Qy 176 LPHNMFVPPVWVFANLSKMWLGGRIRDPILLOSLNEMNTLRTOCGHLYAYDWISIPLV 235
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   316 LGHLKFF-FKAWVDENY-FKAWRQGRGTGYP-LVDAGMR-----LWATGWLH----- 358

Qy 236 YQVVTVAVYGFELTCLVGRFLNPAKAYPGHELDLVVPVTFLOQFFYVGNLKVGLSRA 295
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   359 --DRIRVVVSFFVKV----- 373

Qy 296 LLGMRHGQRGHGQQLLETRMCQERKVRSSVSSOAWNR--TPVIPATREAE-----AG 346

```

Db 374 QLPWRGKMYFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
Db 424 YKFDNGBYVRRWLPESLRLPTDWHHPWNPAPESVLQAGIELG 467

RESULT 4

US-08-272-255-9
; Sequence 9, Application US/08272255
; Patent No. 5824859
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,255
; FILING DATE: 08-JUL-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-272-255-9

Query Match 4.1%; Score 97; DB 2; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

QY 123 LLRRTLYRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFTMPAEHKOLEKLS 175
Db 267 LVRIKQVAVANEAGNEAGESVNLFLKSLGLEYRYSIFNH-----PYSH---ERPL 315
QY 176 LPHNMFVWPVWFANLSKAWLGGRIRDPILLOSLLNEMNTLRQCCHLYAYDWTISIPLV 235
Db 316 LGHLKFF-PAVDENY-FKAWRQRTGY-P-LVDAGMRE-----LWATGWLH- 358
QY 236 YTVVTVVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFPYVGVKVLGSLRA 295
Db 359 --DRIRVVVSFFVKVL----- 373
QY 296 LLGWRHGQRGHGQQLLETRMOCQERKVSRSVSSQAWMR--TPVIPATREAE-----AG 346
Db 374 QLPWRGKMYFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381

Db 424 YKFDNGBYVRRWLPESLRLPTDWHHPWNPAPESVLQAGIELG 467

RESULT 5

PCT-US95-08565-6
; Sequence 6, Application PC/TUS9508565
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08565
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/272,255
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-08565-6

Query Match 4.1%; Score 97; DB 5; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

QY 123 LLRRTLYRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFTMPAEHKOLEKLS 175
Db 267 LVRIKQVAVANEAGNEAGESVNLFLKSLGLEYRYSIFNH-----PYSH---ERPL 315
QY 176 LPHNMFVWPVWFANLSKAWLGGRIRDPILLOSLLNEMNTLRQCCHLYAYDWTISIPLV 235
Db 316 LGHLKFF-PAVDENY-FKAWRQRTGY-P-LVDAGMRE-----LWATGWLH- 358
QY 236 YTVVTVVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFPYVGVKVLGSLRA 295
Db 359 --DRIRVVVSFFVKVL----- 373
QY 296 LLGWRHGQRGHGQQLLETRMOCQERKVSRSVSSQAWMR--TPVIPATREAE-----AG 346
Db 374 QLPWRGKMYFWDITLLD-----ADLESALGQVITGTLPLDSREFRIDNPQEG 423
QY 347 ESLEPGRR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
Db 424 YKFDNGBYVRRWLPESLRLPTDWHHPWNPAPESVLQAGIELG 467

RESULT 6

PCT-US95-08565-9
 ; Sequence 9, Application PC/TUS9508565
 ; GENERAL INFORMATION:
 ; APPLICANT: Greengard, Paul
 ; APPLICANT: Porton, Barbara
 ; APPLICANT: Kao, Hung-Teh
 ; APPLICANT: Ahmad, Margaret
 ; APPLICANT: Lin, Chien-ao
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 ; TITLE OF INVENTION: Using the Same
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
 ; STREET: One Liberty Place, 46th floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/08565
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/272,255
 ; FILING DATE: 08-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary Ph.D., Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: UPN-1795
 ; TELEPHONE: (215) 568-3100
 ; TELEFAX: (215) 568-3439
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 681 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-08565-9

Query Match 4.1%; Score 97; DB 5; Length 681;
 Best Local Similarity 21.5%; Pred. No. 0.23;
 Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;
 QY 123 LLRRLIRYANLGNV-----LILRSVSTAVYKRPSPSAHLVQAGFMTPTAEHKOLEKLS 175
 DB 267 LVRIKQVAVANEENAGEESVNLFLKSLGRLSYRIISFNH-----PYSH-----ERPL 315
 QY 176 LPHNMFVPPVWFWANLSMKAWTGGRIKIRDPILLOSLNENMTLRTQGHLYAYDWTISIPLV 235
 DB 316 LGHLKFP-FWADENY-FKAWRQGRGTGP-LVDAGMRE-----LWATGWLH----- 358
 QY 236 YQVVTVAVYSFPLTCLVGRQFLNPAKAYPGHELDLVVFTFLOFFVVGWLVGLSRA 295
 DB 359 --DRIVVVSFFVKVL----- 373
 QY 296 LLGWRHGQRGHOQLLETRMQCQKRVSRVSSQAWR--TPVIPATREAE-----AG 346
 DB 374 QLPWRGNGKYFDWTLDD-----ADLESALGNGVITGTLDPDSREFRIDNPQEG 423
 QY 347 ESLPGR--RLWQSSSTPLERM-----MMILRPTGLSTG 381
 DB 424 YKFDPNGEYVRWLPELSRLPLTDWIHPWNPAPESVLQAAGIELG 467

RESULT 7
 US-08-906-865-3
 ; Sequence 3, Application US/08906865
 ; Patent No. 6040168

GENERAL INFORMATION:
 ; APPLICANT: Greengard, Paul
 ; APPLICANT: Porton, Barbara
 ; APPLICANT: Kao, Hung-Teh
 ; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David A. Jackson, Esq.
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th
 ; STREET: Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/906,865
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-202
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 582 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; DESCRIPTION: /desc = "Synapsin Ila"
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; US-08-906-865-3

Query Match 4.1%; Score 95.5; DB 3; Length 582;
 Best Local Similarity 20.8%; Pred. No. 0.26;
 Matches 85; Conservative 53; Mismatches 135; Indels 135; Gaps 19;
 QY 110 VSGFVGEKDEGRLLRRLIRYAN---LGNVLILRSVSTAVYKRPSPSAHLVQAGFMTPTA 166
 DB 135 VLGDYDIKVEQAEPSELNLVAHADGTVAVDNQVLN-GTKVRSFRPDPFLIRQHAFGMA 193
 QY 167 EHKOLEKL-----SLP-----HNMFWPPVWFWANLSMKAWLGGRIKIRDPILLOSL 211
 DB 194 ENEDFRHLIIGVAGLPSINSLSIYFCDKPPVFAQLVAIYKTLGGE-KFPLIEQTY 252
 QY 212 ---NEMNTLRT-----OCGHLYA-----YDWISIPLVYQVVTVAVYSFPL----- 249
 DB 253 PNHKEMLTLPFPVVKVIGHAHSCKGVKNHYDFQDIASVVALTQTYATAEFPIDSKY 312
 QY 250 -----TCLVGRQFLNPAKAYPGHELDLVV-----VFTFLQFFVVG 286
 DB 313 DIRVQKIGNNYKATMRTSISGNMKTNTGSAM-----LEQIAMSRYKLWVDTCSEMF---G 365
 QY 287 WLKVLGSLRALLGWRHGQRG-----GQQLLETRMQCQKRVSRVSSQAWW 332
 DB 366 GLDCAVKAV-----HGKDGKDYIFEVNDCSNPLIGEHOVEDRQITELVISKM--NQLLS 419
 QY 333 RTPVIP-----ATREAEAGESLEPGRRLRWQSSSTPLERMMILRPTGLSTGICRCP 387
 DB 420 RTPALSPQRPLTTQPPQSGTLKOP-----DSSKTPPQRP-----PPQG----- 457
 QY 388 WLWNRCTRTCLGWSRTCTGISPSHSPTQLLPPSSVEPPLWAPPSTSA 435

Db 458 -----GPGQPGMPPGKVLPPRLPPGSLPSSSS 489

RESULT 8

US-09-129-668-3
 ; Sequence 3, Application US/09129668B
 ; Patent No. 6429010
 ; GENERAL INFORMATION:
 ; APPLICANT: Greengard, Paul
 ; APPLICANT: Porton, Barbara
 ; APPLICANT: Kao, Hung-Teh
 ; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: 600-1-202 CIP
 ; CURRENT APPLICATION NUMBER: US/09/129,668B
 ; EARLIER FILING DATE: 1998-08-05
 ; EARLIER APPLICATION NUMBER: 08/906,865
 ; EARLIER FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 582
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-129-668-3

Query Match 4.1%; Score 95.5; DB 4; Length 582;
 Best Local Similarity 20.8%; Pred. No. 0.26;
 Matches 85; Conservative 53; Mismatches 135; Indels 135; Gaps 19;
 QY 110 VSGFVEGKDEQRLRLRLIRYAN---LGNVLILRSVSTAVYKRPFSQAHLVQAGFMTPA 166
 Db 135 VLGDYDIKVEQAEPSELNLVAHADGTAYAVDMQVLEN-GTKVRSRPDPVLIRQHAFGWA 193
 QY 167 EHKQLEK-----SLP-----HMFVTPWTFANLSKMWKALGGRIRDPILLOSL 211
 Db 194 ENEDRHLIIGQYAGLSINSLSIYNFCDPKPVFAQVAIYKTLGGE-KPPLIEQTVY 252
 QY 212 ---NEMNTLRT-----QCGLHYA---YDWISIPLYVTQVTVAVYSFPL----- 249
 Db 253 ENHKEMLTLTPFVVVKGHAHSGMGKVENHYDFQDIASVVALTQYVATAPEDISKY 312
 QY 250 -----TCLVGRQFLNPAKAYPGHBLDLPV-----VFTELOFPFVYG 286
 Db 313 DIRVQKIGNNYKAVNRTSISGNWKTNTGSAM-----LEQIAMSRYKLWVDTCSEMF---G 365
 QY 287 WLKVLGSLRALLGWRHGQRGH-----CQQLLETRMQCQERKVSREVSQAWW 332
 Db 366 GLDICAVKAV-----HGKDKDYIFVEMDCSMPLIGEHOVEDRQLITELVISKM--NQLLS 419
 QY 333 RTPVIP-----ATREAEAGESLEPGRRRLWQSSSTPLERMMLIRPTGLSTGICRCPC 387
 Db 420 RTPALSPORLTTQPOQSGLKDP-----DSKTPQRP-----PPQG----- 457
 QY 388 WLWNRCTKCLGNSKTCGISESHSPPTQLPPSSVEPPLWAPSTSA 435
 Db 458 -----GPGQPGMPPGKVLPPRLPPGSLPSSSS 489

RESULT 9

US-08-785-928-1
 ; Sequence 1, Application US/08785928
 ; Patent No. 6087115
 ; GENERAL INFORMATION:
 ; APPLICANT: Gerhengorn, Marvin C.
 ; APPLICANT: Arvanitakis, Leandros
 ; APPLICANT: Geras-Raaka, Elizabeth
 ; APPLICANT: Cesarman, Ethel
 ; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED
 ; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
 ; TITLE OF INVENTION: ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
 ; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 ; STREET: Clinton Square, P.O. Box 1051
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 14603

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/785,928
 ; FILING DATE: 22-JAN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Goldman, Michael L.
 ; REGISTRATION NUMBER: 30,727
 ; REFERENCE/DOCKET NUMBER: 19603/1320
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (716) 263-1304
 ; TELEFAX: (716) 263-1600
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 342 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-785-928-1

Query Match 4.0%; Score 95; DB 3; Length 342;
 Best Local Similarity 24.6%; Pred. No. 0.14;
 Matches 69; Conservative 41; Mismatches 110; Indels 60; Gaps 15;
 QY 3 ITYTSQVANARLGSFRLL--CWRGSIYKL-LYGEFLIFLLCYVIRFIYRLALTEQQ 59
 Db 68 VTIYFKHRSRAGADIDILLGICLNSCLSLSLAEVLMFLFPNIISTGLCEL-----E 121
 QY 60 LMPEKLTLYCDSYQIDIPISFVLGVTVLVTRWNNQYENLPWDRMLSLVSGFVEGKD- 118
 Db 122 IFFYLYVYLDIF-SVVCVSLV-RYLLVAYSTRSNPKKQSLGWLTSAAALLIALVLSGDA 179
 QY 119 --EGRCL---LRRTLIRYANLGNV-----LILRSYSTAVYKRPFSQAHLVQAGFMTPAEH 168
 Db 180 CRHSRVVDPSVKQAMCYENAGNMTADWRLHVRIVS-----VTAGFLLP--- 223
 QY 169 KOLEKLSLPHNMFVPPWFWFANLSKMWKALGGRIRDPILL-----QSLNEMNTL--- 217
 Db 224 --LALLILFYALTWCV-VRRTKLQARRKVRGVIVAVVLLFFVFCFPYHVLNLLDLLRRR 280
 QY 218 --RTQCGLHYADWISIPLVYT---QVTVVAVYSFTECL 252
 Db 281 WIRDSC---YTRGLINVLAVTSLLOALYSAAVPLIYSCL 317

RESULT 10

US-08-728-603-17
 ; Sequence 17, Application US/08728603
 ; Patent No. 6093806
 ; GENERAL INFORMATION:
 ; APPLICANT: Cesarman, Ethel
 ; APPLICANT: Knowles, Daniel M.
 ; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
 ; TITLE OF INVENTION: HERPESVIRUS
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
 ; STREET: Clinton Square, P.O. Box 1051
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: USA

ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US/08/728,603
ATTORNEY/AGENT INFORMATION:
NAME: BRAUN, SUSAN J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19603/720
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1636
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-603-17

Query Match 4.0%; Score 95; DB 3; Length 342;
Best Local Similarity 24.6%; Pred. No. 0.14;
Matches 69; Conservative 41; Mismatches 110; Indels 60; Gaps 15;

QY 3 IYTSQVANARIGSFRLL--CWGRSIIYL-LYGEFLIFLCYIIRFIYRLALTEQO 59
DB 68 VTYIFCKHRSRAGADILLGLICLSLSLAEVLMFLPNIISTGLCLL-----E 121
QY 60 LMFEKLTLYCDYIQLIPISFVLGVVTLVTRWNOYENLWPDRLMSLVSGFVEGKD- 118
DB 122 IFYYLYVLDLIP-SVVCVSLV-RYLLVAYSTRSNPKOSGLVLTSAALLIALVLSGDA 179
QY 119 --EOGRL--LRRTLIYRANLGNV-----LILRSVSTAVYKFPSPSAQHLVQAGFWTPASH 168
DB 180 CHRHRVDDPVSKQAMCYENAGNMTADWRLHRTVS-----VTAGFLLP---- 223
QY 169 KOLEKLSLPHNMFVWVWFANLSKMWLGGRIRDPILL-----QSLLNEMNTL- 217
DB 224 --LALLILFYALTWCV-VRRTKQARRKVRGIVAVVLLFFVFCPPHYVNLNLLTLRRR 280
QY 218 --RTOGCHLYAYDWISIPLVYT--QVVTVAVYSFELTCL 252
DB 281 WIRDSC---YTRGLINVLGAVTSLIQALYSVAVVLIYSCL 317

RESULT 11
US-08-964-268-5
Sequence 5, Application US/08964268
Patent No. 6114503
GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: RUBEN, STEVEN M
APPLICANT: SANCAR, AZIZ
APPLICANT: HSU, SHINO-WEN D
APPLICANT: KAZANTSEV, ALEKSEY G
TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,268
FILING DATE: HEREWITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,189
FILING DATE: 04-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 681 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-964-268-5

Query Match 4.0%; Score 94; DB 3; Length 681;
Best Local Similarity 21.5%; Pred. No. 0.47;
Matches 61; Conservative 29; Mismatches 86; Indels 108; Gaps 14;

QY 123 LLRRTLIYRANLGNV-----LILRSVSTAVYKFPSPSAQHLVQAGFWTPASHKOLEKLS 175
DB 267 LVRIKQVAVANEGNEAGEESVNLFLKSLGIREYSYISFNH-----PYSH--ERPL 315
QY 176 LPHNMFVWVWFANLSKMWLGGRIRDPILLQSLNEMNTLRTQCHLYAYDWISIPLV 235
DB 316 LGHLKFF-PAVDENY-FRAMRQGRGTYP-LVDAGMRE-----LWATLWLH----- 358
QY 236 YQVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFVYVGLKVLGSLRA 295
DB 359 --DRIRVVVSFFVKVL----- 373
QY 296 LLGWRHGORGHQQLLETRMQCOEKVRVRESSQAWR--TPVIPATREAE-----AG 346
DB 374 QLPWRGMKYPWDTLLD-----ADLESDALGMQYITGTLDPDSREFRIDNPQFEG 423
QY 347 ESLEPGR--RLMWQSSSTPLERM-----MMILRPTGLSTG 381
DB 424 YKFDNGEYVRWLPPELSRLPTDWHHPWNAPESVLQAGIELG 467

RESULT 12
US-08-229-418-2
Sequence 2, Application US/08229418
Patent No. 5712111
GENERAL INFORMATION:
APPLICANT: Linemeyer, David L.
APPLICANT: Menke, John G.
APPLICANT: Hess, John F.
APPLICANT: Borkowski, Joseph A.
APPLICANT: Bierillo, Kathleen K.
TITLE OF INVENTION: DNA ENCODING BRADYKININ B1 RECEPTOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen III
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

STATE: New Jersey
 COUNTRY: USA
 ZIP: 07065-0907
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/307,912
 FILING DATE: 10-May-1999
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Hand, J. Mark
 REGISTRATION NUMBER: 36,545
 REFERENCE/DOCKET NUMBER: 19202CA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (732) 594-3905
 TELEFAX: (732) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 353 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-307-912-2

Query Match 3.9%; Score 92.5; DB 4; Length 353;
 Best Local Similarity 19.6%; Pred. No. 0.26;
 Matches 74; Conservative 45; Mismatches 108; Indels 151; Gaps 18;

QY	38	IFLLCYII---RFIYRLALTEBQQLMPEKLTLYCDSYIQLIPISFVLG--FYVTLVVTR	92
Db	44	IISICFFGLGNLFLVLLPRQLNVAEIVL---ANLAASDLVFLGLPFAENI---	97
QY	93	WNOYENLPW-----DRLMSLVSGFVEGKDEQGR	123
Db	98	WNOFN---WPFGLLCRVINGVIKANLFIISFLVVAISQDRYRVLHPMASGRQRRQ	153
QY	124	LRRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHMFW	183
Db	154	ARVTCVLIVVVGGL---SIPTFLRSIQAVPDNITACI-----LLLPH---	195
QY	184	PWVFNLSMKAWLGGRIID-----PI-----LQSLNEMNTLRTQCHLYA	226
Db	196	-----EAMHPARIVELNIGLFLPLAAIVFFNVHILASLRTREVSRTCGG--R	243
QY	227	YDWISIPLVYTVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFVY-	285
Db	244	KDSKTTALILTLLVA-----FLVCWAPYHF-----PAPLEFLFQVQ	279
QY	286	-----GWLKVGLSRA-----LLGWRHGQRGHGOQLLETRMOCQERK	321
Db	280	AVRGCFWEDFDLGLQLANFAFTNSLNPNVIYFVG-----RLFRKWNELYKQCTPKS	334
QY	322	VSREVS-----QAWR	333
Db	335	LAPISSSHRKEIFQLFWR	352

RESULT 15

PCT-US95-04464-2
 Sequence 2, Application PC/TUS9504464
 GENERAL INFORMATION:
 APPLICANT: Linemeyer, David L.
 APPLICANT: Menke, John G.
 APPLICANT: Hees, John P.
 APPLICANT: Borkowski, Joseph A.
 APPLICANT: Bierillo, Kathleen K.
 TITLE OF INVENTION: DNA ENCODING BRADYKININ B1 RECEPTOR
 NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
 ADDRESSEE: John W. Wallen III
 STREET: P.O. Box 2000
 CITY: Rahway
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07065
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04464
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Wallen III, John W.
 REGISTRATION NUMBER: 35,403
 REFERENCE/DOCKET NUMBER: 19202
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-3905
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 353 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04464-2

Query Match 3.9%; Score 92.5; DB 5; Length 353;
 Best Local Similarity 19.6%; Pred. No. 0.26;
 Matches 74; Conservative 45; Mismatches 108; Indels 151; Gaps 18;

QY	38	IFLLCYII---RFIYRLALTEBQQLMPEKLTLYCDSYIQLIPISFVLG--FYVTLVVTR	92
Db	44	IISICFFGLGNLFLVLLPRQLNVAEIVL---ANLAASDLVFLGLPFAENI---	97
QY	93	WNOYENLPW-----DRLMSLVSGFVEGKDEQGR	123
Db	98	WNOFN---WPFGLLCRVINGVIKANLFIISFLVVAISQDRYRVLHPMASGRQRRQ	153
QY	124	LRRLIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHMFW	183
Db	154	ARVTCVLIVVVGGL---SIPTFLRSIQAVPDNITACI-----LLLPH---	195
QY	184	PWVFNLSMKAWLGGRIID-----PI-----LQSLNEMNTLRTQCHLYA	226
Db	196	-----EAMHPARIVELNIGLFLPLAAIVFFNVHILASLRTREVSRTCGG--R	243
QY	227	YDWISIPLVYTVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFVY-	285
Db	244	KDSKTTALILTLLVA-----FLVCWAPYHF-----PAPLEFLFQVQ	279
QY	286	-----GWLKVGLSRA-----LLGWRHGQRGHGOQLLETRMOCQERK	321
Db	280	AVRGCFWEDFDLGLQLANFAFTNSLNPNVIYFVG-----RLFRKWNELYKQCTPKS	334
QY	322	VSREVS-----QAWR	333
Db	335	LAPISSSHRKEIFQLFWR	352

Search completed: July 10, 2003, 12:27:53
 Job time : 10.9682 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: July 10, 2003, 12:22:04 ; Search time 17.1292 Seconds
(without alignments)
3283.199 Million cell updates/sec

Title: US-09-622-964-3
Perfect score: 3120
Sequence: 1 MTITYTSQVANARLGSFRL.....TLKDHDPYWALENRDEAHS 585
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	814	26.1	612	2 T32368	hypothetical prote
2	770.5	24.7	557	2 T32367	hypothetical prote
3	758	24.3	584	2 T19565	hypothetical prote
4	752	24.1	632	2 S44917	zk688.2 protein -
5	751	24.1	405	2 T27971	hypothetical prote
6	717.5	23.0	1355	2 T28715	hypothetical prote
7	717	23.0	499	2 T27630	hypothetical prote
8	682	21.9	413	2 T21644	hypothetical prote
9	679	21.8	450	2 T18781	hypothetical prote
10	669.5	21.5	523	2 T18782	hypothetical prote
11	656	21.0	459	2 S40708	hypothetical prote
12	646.5	20.7	513	2 T24210	hypothetical prote
13	632.5	20.3	405	2 S42371	hypothetical prote
14	622.5	20.0	613	2 T16885	hypothetical prote
15	622	19.9	420	2 B89710	protein C43G2.4 [i
16	616.5	19.8	512	2 T19806	hypothetical prote
17	585	18.8	400	2 T21670	hypothetical prote
18	583	18.7	387	2 H89192	protein F32G8.4 [i
19	495	15.9	400	2 T20050	hypothetical prote
20	470	15.1	530	2 T28037	hypothetical prote
21	455	14.6	444	2 T20048	hypothetical prote
22	447	14.3	806	2 T15468	hypothetical prote
23	441.5	14.2	411	2 T28038	hypothetical prote
24	395	12.7	434	2 T20922	hypothetical prote
25	127.5	4.1	315	2 AB0677	probable membrane
26	115.5	3.7	613	2 B82338	conserved hypothet
27	115	3.7	516	2 AE0665	probable membrane
28	113	3.6	932	2 S62555	protoplast regener
29	111.5	3.6	825	2 T47164	hypothetical prote

30	111	3.6	481	2 T42226	hypothetical prote
31	107	3.4	2606	2 T03159	large tegument pro
32	105	3.4	1707	2 T18951	hypothetical prote
33	104.5	3.3	457	2 D71505	probable transloca
34	104.5	3.3	457	2 S44484	secy protein - chl
35	104.5	3.3	785	2 A86299	hypothetical prote
36	104.5	3.3	786	2 H64817	probable membrane
37	104	3.3	485	2 T38255	hypothetical prote
38	103.5	3.3	741	2 F90739	probable transport
39	103.5	3.3	741	2 H85589	probable transport
40	103	3.3	321	2 G90894	hypothetical prote
41	103	3.3	321	2 B85723	hypothetical prote
42	103	3.3	961	2 S58787	UFD2 protein - yea
43	102.5	3.3	321	2 C64906	probable membrane
44	102.5	3.3	1002	2 T43236	carboxypeptidase C
45	101.5	3.3	343	2 T40306	hypothetical prote

ALIGNMENTS

RESULT 1

T32368
hypothetical protein C01B12.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C:Accession: T32368
R:Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid C01B12.
A:Reference number: Z21156
A:Accession: T32368
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-612 <SCH>
A:Cross-references: EMBL:AF025458; PIDN:AAB70976.1; GSPDB:GN00020; CESP:C01B12.3
A:Experimental source: strain Bristol N2; clone C01B12
C:Genetics:
A:Gene: CESP:C01B12.3
A:Map position: 2
A:Introns: 25/3; 60/2; 105/2; 138/3; 212/3; 319/3; 369/2; 467/2; 508/3; 573/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match		26.1%	Score 814;	DB 2;	Length 612;
Best Local Similarity		33.6%	Pred. No. 2.3e-55;		
Matches 193;		Conservative 98;	Mismatches 221;	Indels 62;	Gaps 10;
QY	1	MTITYTSQVANARLGSFRLLLCWGSGIYKLYGFLFLPCLCYIIRFYRLALTEEQQL	60		
DB	1	MTVTYSLDVASSSFCLYKLLFRMKGSIWKSWAELVWVWLCYAVLSVIYRCLLTWKQRA	60		
QY	61	MFEKLTLYCDSVIQLIPISFVLGFFVTVLVTRWVQENLWPDRLMSLVSGFVSGKDFQ	120		
DB	61	TFEDLCIFDFYSNIPITFMLGFYSAVFRWQIFDNIGWIDTFCLWITQYIKETER	120		
QY	121	SRLRLRTIRYANLGNVLILRSVSTAVYKRPSSAHLVQAGPMTAEHKLEKLSLPHNM	180		
DB	121	AKCVRENCIRYSILTQANVYRDVAASVKRPTFNHLVTAGLMTKEKAEFESISPHAK	180		
QY	181	FWPVMWPFANISMAKWLGGRIIDPILLOSLLNEMNTLTQCCHLYAYDWISIPLYVTQV	240		
DB	181	YWQPMHMLFSMITLARDEGMISDIIYVDLMKQRQRVRNLTSLTDFDWPVPLVYTVQV	240		
QY	241	TVAVYSPLTCLVGRQFENLPAKAYPG---HELDLVVVFTELOFFFYVWLNKVAQLNP	297		
DB	241	HLAVRSYELIALFGRLHPESNRINDPKQITDLYVPIMSLLQFFITFGWVKVAEVLNP	300		
QY	298	FEDEDDPFTNWIVDRLNQVSLAVDENHQDLPRMPEPMYWNKPPQPPYTAASQFRRA	357		
DB	301	LGEDDDDFECNWLDRNLQVGLWVVDYATNYPTLEKQDFWEDAEALYTAESAMRPLN	360		
QY	358	SFMGSTFNISLUNKEMEFPQNOE-----DEDAHAGIIGFLGQ-----	397		

Db 361 PQVSCADMPTEBPFMVPRRTLSRMSHWGDMEDTD---VVVVVGLKXTRDONSVAS 417
QY 398 -----SHDHPFRANSRTKLLWPKESSLHLEGLPKNHKAQKQVRGOE 440
Db 418 GESLAFNSFANGGRKLESEMPFRAGSRIGDRYKRNSSAQDP--ENGMAKQNSIDENA 475
QY 441 DNKAWKL-KAVDAPKSGPLVQPGYSAQPPFLSP-----TPMPFPLEPSAPSKLHS 491
Db 476 DIHSNRLDQASGTPKSGRLWS-----SMPQTLEMLKNKFNFPVYNTDGMKDRLOQN 530
QY 492 VTGI-DTKDKSLKTVSSGA--KKSFELLESDDGA 522
Db 531 PTPITDHIDPLHVASSQSWMFNLSPVKEEEA 564

RESULT 2

T32367
hypothetical protein C01B12.5 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C/Accession: T32367
R/Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A/Description: The sequence of C. elegans cosmid C01B12.
A/Reference number: Z21156
A/Accession: T32367
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-557 <S>
A/Cross-references: EMBL:AF025458; PIDN:AB70978.1; GSPDB:GN00020; CESP:C01B12.5
A/Experimental source: strain Bristol N2; clone C01B12
C/Genetics:
A/Gene: CESP:C01B12.5
A/Map position: 2
A/Introns: 60/2; 105/2; 212/3; 318/3; 368/2; 442/1; 509/3
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 24.7%; Score 770.5; DB 2; Length 557;
Best Local Similarity 34.3%; Pred. No. 5e-52;
Matches 162; Conservative 90; Mismatches 171; Indels 49; Gaps 6;
QY 1 MTITYTQVAVNARLGSRLLCWRSYKLLYGEPLIFLLCYIIRFYRLALTEQQQL 60
Db 1 MTVAISLDVATSGELTQIKVLLRWKSGVMSIWSSELLIWLCCYSILSVIRILLNKAQRE 60
QY 61 MPEKLYCDSTYQILPISFVLGFYTLVVTRWQVNLNPPDRLMSLVSGFVEGKDEQ 120
Db 61 VFEQLCTFDFTSFVPIVTFMLGFYSIVYNRKTVFDNVGMDTSALTIAQVIRTSK 120
QY 121 SLLRLTLRYANLGNVLIIRSVAVYKRPQSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180
Db 121 ARLIRENCVRYMVAQVMVFRDVSPIARRRFPPIKHLIGAGLLTDELTEFDAITSQSK 180
QY 181 FWVPVW 240
Db 181 YWQPIQWLSLVTVAKDEGLIADSYLVLDIKQREFRILKMLNLFDMVPIPLVTVQV 240
QY 241 TVAVVSFFUTCLVGRFLNPAKAYPG--HELDLVVPVFTFLQFFYVGVKLVKVAEQLNPF 298
Db 241 NLAVRTYFVLAFGRQFLENNNNIPGAKWKIDYFIMTSLQIVFVGVKLVSEVWMLNPL 300
QY 299 GEDDDDFETNWIQVNLQVSLAVDEMHDQLPRMEPDWYKPEPPYTAASAQPRRAS 358
Db 301 GEDDEDFETNWIERNLQVAVVQAYGRYPTLKRDPFWEDETPQMDTPTSTRKAHSH 360
QY 359 FMGSTFNISLNKEEMFQPNQDEEDAHAGII-----GRFLG---LQSHDHPFRANSR 409
Db 361 MQSGSCINM-----NEADLDNGLSYVRRRSRSGFDDDDASSFSTENTKTQSN 406
QY 410 TKLL-----W-----PKRESLHLEGLPKNHKAQKQVR 437
Db 407 ASMLPRHMQQPRNKISAAIAKFKYSEAPRRAVCVVAALNLNHRPSTDLR 458

RESULT 3

T19565
hypothetical protein C29F4.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/Accession: T19565
R/Kershaw, J.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z19143
A/Accession: T19565
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-584 <WIL>
A/Cross-references: EMBL:Z68335; PIDN:CAA92730.1; GSPDB:GN00022; CESP:C29F4.2
A/Experimental source: clone C29F4
C/Genetics:
A/Gene: CESP:C29F4.2
A/Map position: 4
A/Introns: 36/2; 78/2; 104/2; 149/2; 174/2; 221/2; 256/3; 333/3; 412/2; 484/2; 526/1; 5
C/Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match

Best Local Similarity 24.3%; Score 758; DB 2; Length 584;
Matches 172; Conservative 110; Mismatches 195; Indels 64; Gaps 11;
QY 19 RLLLCWRGSIYKLLYGEPLIFLLCYIIRFYRLALTEQQQLMPEKLYCDSYIQLIPI 78
Db 63 KLIFKVKGSLWKAIFYLDLIVMCFCYAFISVYRYALDRSQDTPFERFMQFCNRRLDYPI 122
QY 79 SFVLGFYTLVVTRWQVNLNPPDRLMSLVSGFVEGKDEQSLRLRRLIRYANLGNVL 138
Db 123 NFMLGFYTLVVTRWQVNLNPPDRLMSLVSGFVEGKDEQSLRLRRLIRYANLGNVL 182
QY 139 IIRSVAVYKRPQSAQHLVQAGFMTPAEHKQLEKLSLPHNMFWVWVWVWVWVWVWVW 198
Db 183 VFRDIHIGVRKFPFTEMTVAAGIMTSSELKKYNEVESRYAKYTWLGFNNTFNLLNEARRE 242
QY 199 GRIRDPIQLLQSLNEMNTLRTQCHLYAYDWISIPLVYTVQVTVAVYSFELTCLVGRQFL 258
Db 243 GRIESAYTQNAIAEIRTPRGLSLIWTYDWVPIPLMYPQLVPMATHCYLVCLVSRQFV 302
QY 259 NPAKAYPGHELDLVVPVFTFLQFFYVGVKLVKVAEQLNPFDDDDDFETNWIQVNRLOVS 318
Db 303 INSDAVNTTEIDLGVPEFTIIEFIFVGMVWLVKVAEQLNPFDDDDDFETNWIQVNRLOVS 362
QY 319 LLAVDEMHDQLPRMEPDWYKPEPPYTAASAQPRRASFMGSTFN-----ISLNKEEME 374
Db 363 MGIVDDSHDDGPILEKQMFVN-DTVSPLYSSAAQORNVNFFYFGSATNADAQIPDDVVRQIT 421
QY 375 FQPNQDEEDAHAGIIIGRFLGQSHDHPFRANSRTKLLWPKESSLHLEGLPKNHKAQK 434
Db 422 MIPHPLNEK-----LDQMYGKRT--NRPP-VQSVVEL---KRDQRFSTG-----459
QY 435 NVRGEDNKAWKLKAVDAPKSGPLVQPGYSAQPPFLSPMPFPLEPSAPSKLHSVTG 494
Db 460 -----NNRKQLEAKFNSKLGAMFQKRKSKSLTFTPEST-----KAPASTD 501
QY 495 IDTKDKSLKTVSSGAKKSFELLESDDGALMEHPEVQVQRKRTVEFNLTDPMPDENHLKE 554
Db 502 IET-----VSTFENTOKPCY-----SNPDCIVEVDEESQ-----DIPKVTBEPKQE 542
QY 555 P 555
Db 543 P 543

RESULT 4

S44917
ZK688.2 protein - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C/Accession: S44917

Db 1027 QVFLAVRIYFFLCIARQSVLIDGKPKDQNSPVYP-----FVPFLMTSLQFVYVYVGM 1080
QY 289 KVAEQLINPFGEEDDDDDPETNWIIVDRNLQVSLIAVDEMHQDLPRMEPDYMNKPEPOPPYT 348
Db 1081 KVAESLWNLPGEDDDDDPECNLYLDNRNLAVGLAIVDS--SEAPHYEKDVFGLAIADPLYS 1139
QY 349 AASAQFRASFGWGSTFNISLNGEMEFQPNQ--DEDAHAGIIG-----RFLGLQSH 399
Db 1140 SDTANTHINPQIGSAATYETQENIIRPHVDNDMEDGDDVEGCNPRQLSRFSVSVSV 1199
QY 400 DHPPRANRSTKLLWPKRESLLHGLPKNHKAQKQVNGQEDNKAWLKAVDAFPKSGPLY 459
Db 1200 NRN---CESRTSL--SRNPFTIIRLSQFGSKVN-----1231
QY 460 QRGYYSAPOTPLSPMPFPLPSPKLSHVTGIDTKSLKTSYSSGAKKSFEILLS 519
Db 1232 -RPG-----KLFSSQFSINTANGDN-----DIGSCASILGELAES 1267
QY 520 D-----CALMEHPVSVQRKTVFENLTDMEIPEINHLKBPLEQSPNHTTLKDHMDPTW 575
Db 1268 NKASQGLLTPEYAGSFRNTPIDMLTSVPE-----EDEEAQKTRTSVDLRKWKEMVE 1320
QY 576 ALNRDEA 583
Db 1321 NEKKREEA 1328

RESULT 7
T27630
hypothetical protein ZC518.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
R:Thomas, K.
submitted to the EMBL Data Library, January 1996
A:Accession: T27630
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-499 <WIL>
A:Cross-references: EMBL:Z68753; PIDN:CAA92989.1; GSPDB:GN00022; CESP:ZC518.1
A:Experimental source: clone ZC518
C:Genetics:
A:Gene: CESP:ZC518.1
A:Map position: 4
A:Introns: 30/3; 60/2; 106/2; 239/3; 267/3; 315/3; 329/3; 380/2; 405/3; 445/2; 466/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C01B12.5

Query Match 23.0%; Score 717; DB 2; Length 499;
Best Local Similarity 35.2%; Pred No. 6.3e-48;
Matches 160; Conservative 78; Mismatches 158; Indels 58; Gaps 9;

QY 1 MTITVTSQVANRLGSRLLCWRSYKLYGFLFLCYIIRFYRLALTEEQOL 60
Db 1 MTISYTLDSQTNLQSPFSLLRWGSVWVAFGLAVMTAVFLLSICIYRYMLSPSQD 60
QY 61 MFEKLTLYCDSYIQL-IPISFVLGVFVTLVTRWQYENLWPPRLMSLSGFEVGEKDE 119
Db 61 VFEQIRIFDNKLDANIPLTLFGFVFWVARWGSILNGIWDASLLFATYIRGADE 120
QY 120 QSRLRRTLIIRYANLGNVLIILRSVAVYKRPSPSAQHLVQAGFMTPAEHKOLSLPHN 179
Db 121 ETRVIRRNLYVLVSQALVLRDISQVYKRPFTMTDLAASGLMTHHEWMDLIDHKDYS 180
QY 180 MFWVPVWVFWANLGMKAGGRIRDPILLOSLNEMNTLRTQCGHLYAYDWISIPLVYQV 239
Db 181 RYMTSIQWLSNLVVEQCKGKGVDSYLLMNKIYDEIGKPRHGLASLLKYDWVPEVLVYPOV 240
QY 240 VTVAVYSFELCLVGRQFL-----NPAXAYPGHELDLVVFTFLQFPFYVGLKVAEQLI 295
Db 241 IFLAVRIYFMICLGRQFIVTGPNS-----GIDLWLPITTMVQFVLYMGWKMVAEALL 294
QY 296 NPFGEEDDDDDPETNWIIVDRNLQVSLIAVDEM--HQD-----LPRMEPDYMNKPN 341

Db 295 NPLGEDDDDDLECNYYIIDKNLITGLSIVDTWKKHDDTGYSWVEHMAKTPAQKDBFWGID 354
QY 342 EPQPPYTAASAQFRASFGWGSTFNISL--NKEMEFQPN-----NQ 379
Db 355 KIAPLYSMESABRSVHPLVGSASKINLVKNKKEIVMTPHKNKLSLDPSEQKTYLRRVNV 414
QY 380 EDEDAHAGIIGRFLGLQSHDHPHPRANSRTKLL 413
Db 415 SDHNAKHA---KQKGL-----RANSPDKCL 437

RESULT 8
T21644
hypothetical protein F32B6.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21644
R:Basham, V.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19453
A:Accession: T21644
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <WIL>
A:Cross-references: EMBL:Z81074; PIDN:CAB03043.1; GSPDB:GN00022; CESP:F32B6.9
A:Experimental source: clone F32B6
C:Genetics:
A:Gene: CESP:F32B6.9
A:Map position: 4
A:Introns: 22/2; 61/1; 141/2; 176/3; 200/3; 290/3; 332/1

Query Match 21.9%; Score 682; DB 2; Length 413;
Best Local Similarity 39.6%; Pred No. 2.6e-45;
Matches 141; Conservative 67; Mismatches 126; Indels 22; Gaps 7;

QY 1 MTITVTSQVANRLGSRLLCWRSYKLYGFLFLCYIIRFYRLAL-----55
Db 1 MTISYS-----GNVIRILLRWKGSIRWTAKKELLIVLYSVRVFLKIGIDLDD 52
QY 56 EQQL-----MFEKLTLYCDSYIQLIPISFVLGVFVTLVTRWQYENLWPPDRMLSLVS 111
Db 53 EDDLKWRMEFETPCRCQDSYRLIPLTLFLGFSVSNVAVRWQRQFETLYWPEDLSVL 112
QY 112 GFVEGKDEQSRLRRTLIIRYANLGNVLIILRSVAVYKRPSPSAQHLVQAGFMTPAEHKOL 171
Db 113 TVLHQHDEKSKRRRTTIARYLNLANALAWRDISSKIRLRFPSVHSLIESGLLTKYQIL 172
QY 172 EKLSLPH-NMFWVPVWVFWANLGMKAGGRIRDPILLOSLNEMNTLRTQCGHLYAYDWI 230
Db 173 EAMAEANESSRWITPLHWIQLIMRQVBEHKKPTASLNFQFVGLRIFRQSLRKLYSYDW 232
QY 231 SIPLVYTVQVTVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVFTFLQFPFYVGLK 290
Db 233 CVPLVYTVQVVALATYSFPFPLFGQFLFP-DITGKELDLVVPVFTIVQVFLFVGFVK 291
QY 291 AEQLINPFGEDDDDDPETNWIIVDRNLQVSLIAVDEMHDQ-LPRME--PDYMNKPEP 343
Db 292 QDLMRPFGLDDDDIELNYILDRNVRISFAIVNQLESPIPDFESNDDKLWHEHP 347

RESULT 9
T18781
hypothetical protein B0564.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T18781
R:Lightning, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19021
A:Accession: T18781
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-450 <WIL>
A:Cross-references: EMBL:Z73422; PIDN:CAA97765.1; GSPDB:GN00022; CESP:B0564.3
A:Experimental source: clone B0564
C:Genetics:
A:Gene: CESP:B0564.3
A:Map position: 4
A:Introns: 60/2; 105/2; 320/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.8%; Score 679; DB 2; Length 450;
Best Local Similarity 32.9%; Pred. No. 5e-45;
Matches 155; Conservative 80; Mismatches 184; Indels 52; Gaps 9;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLLYGEBFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINYHKEIKTSHTWKFVLLPRWKGSIWKAIIYMETIIPILCYIGIISVYVVTAMSEBSQR 120
QY 61 MPEKLTLYCDSYIQLPIPSIFVLGFFYTLVTRWNOYENLPWDRMLSLVSGFVEGKDEQ 120
DB 61 TFSVIRYCDKGLSPFLPFLVGLFFVTVVDRWTKLWRTVGFDDVCLLANLVVGTSEK 120
QY 121 SRLRLTLIRYANLGNVLILRSYSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHN- 179
DB 121 AIYRENIARYCALTOLLVFRDYSMTRRFPMTETVVAAGFMSKDELIDYNSYTTKNS 180
QY 180 ----MFWVPWVWFANLWGLGGRIRDPILLQSLNEMNTLRTQCGHLYAYDWISIPLV 235
DB 181 RLKGYKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIEWCNYDVPPLM 240
QY 236 YQVTVVAVSYFELCLVGRQFLNPAKAYPGHLDLVVPVTFPQFFYVGVWGLKVAEQLI 295
DB 241 YPOLVCLAVNLVFLVSIIRQ-LVTEKHQKQVDEVDFVFFVMTFLOFIFYMGWLKVEVLM 299
QY 296 NPFGEDEDDFETNWIIVDRNLQVSLAVDEMHDLPMEPDYMNKPEPQPPYTAASAOQR 355
DB 300 NPFGEDEDDFETNALIDRNITWGLKXVONT-MKTPELLKQDFDEVLVLLYSESSQIS 358
QY 356 RASFNGSTFNISLNK--EMEPQNOEDE 382
DB 359 NYHYHGSTSEVHLEQKCSVRMIPHSQE 387

RESULT 11
S40708
hypothetical protein C07A9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Aug-2000
C:Accession: S40708
R:Smith, M.
submitted to the EMBL Data Library, December 1993
A:Reference number: S40701
A:Accession: S40708
A:Molecule type: DNA
A:Residues: 1-459 <SMI>
A:Cross-references: EMBL:Z29094; NID:G436440; PID:G436448
C:Genetics:
A:Introns: 40/3; 77/2; 103/2; 109/2; 154/2; 179/2; 261/3; 411/2
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.0%; Score 656; DB 2; Length 459;
Best Local Similarity 32.3%; Pred. No. 3.2e-43;
Matches 129; Conservative 91; Mismatches 151; Indels 28; Gaps 5;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLLYGEBFLIFLLCYIIRFYRLALTEEQ- 59
DB 44 LSYNVYDLATSKSLMIVRMIFKWRGSLWQAVYKELIVWICAYSLVSVIYRFPALTRSQE 103
QY 60 ----LMPEKLTLYCDSYIQLPIPSIFVLGFFYTLVTRWNOYENLPWDRMLSLVSGFV 114
DB 104 QNKBIIIIFERFEGYCDARMGYLPNLFVGLFFCNIIRRLKLYTSLGIDNIALFVSAYV 163
QY 115 EGKDEQSRLRLTLIRYANLGNVLILRSYSTAVYKRFPSAQHLVQAGFMTPAEHKQLEK 174
DB 164 RGTDRARQRNRNIIRYCVISQCLVFRDIHVGVRRRPTLEAVAQAGIMLPHLEKFNIS 223
QY 175 SLPHNFWVWVWFANLWGLGGRIRDPILLQSLNEMNTLRTQCGHLYAYDWISIP 234
DB 224 KSRYKQYVWVWFANLWGLGGRIRDPILLQSLNEMNTLRTQCGHLYAYDWISIP 283
QY 235 YQVTVVAVSYFELCLVGRQFLNPAKAYPGHLDLVVPVTFPQFFYVGVWGLKVAEQ 294
DB 284 MYPOLVNMVAVHTYFELCLVGRQFLNPAKAYPGHLDLVVPVTFPQFFYVGVWGLKVAEQ 343
QY 295 INPGEDEDDFETNWIIVDRNLQVSLAVDEMHDLPMEPDYMNKPEPQPPYTAASAOQR 346
DB 344 LNFGEDEDDFETNWIIVDRNLQVSLAVDEMHDLPMEPDYMNKPEPQPPYTAASAOQR 403
QY 347 -YTAASAOQRFRASFMGSGTFTNLSNKEEMEF-----QPNQ 379

RESULT 10
T18782
hypothetical protein B0564.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T18782
R:Lightning, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19021
A:Accession: T18782
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-523 <WIL>
A:Cross-references: EMBL:Z73422; PIDN:CAA97766.1; GSPDB:GN00022; CESP:B0564.4
A:Experimental source: clone B0564
C:Genetics:
A:Gene: CESP:B0564.4
A:Map position: 4
A:Introns: 60/2; 105/2; 320/3; 405/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 21.5%; Score 669.5; DB 2; Length 523;
Best Local Similarity 36.5%; Pred. No. 3.4e-44;
Matches 142; Conservative 73; Mismatches 165; Indels 9; Gaps 4;

QY 1 MTITYTSQVANARLGSRLLLCWRGSIYKLLYGEBFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINYHKEIKTSHTWKFVLLPRWKGSIWKAIIYMETIIPILCYIGIISVYVVTAMSEBSQR 120
QY 61 MPEKLTLYCDSYIQLPIPSIFVLGFFYTLVTRWNOYENLPWDRMLSLVSGFVEGKDEQ 120
DB 61 TFSVIRYCDKGLSPFLPFLVGLFFVTVVDRWTKLWRTVGFDDVCLLANLVVGTSEK 120
QY 121 SRLRLTLIRYANLGNVLILRSYSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHN- 179
DB 121 AIYRENIARYCALTOLLVFRDYSMTRRFPMTETVVAAGFMSKDELIDYNSYTTKNS 180
QY 180 ----MFWVPWVWFANLWGLGGRIRDPILLQSLNEMNTLRTQCGHLYAYDWISIPLV 235
DB 181 RLKGYKIWIIPANWALCMYKARKDGYIESDYFKAQMEGEIRTWRTNIEWCNYDVPPLM 240
QY 236 YQVTVVAVSYFELCLVGRQFLNPAKAYPGHLDLVVPVTFPQFFYVGVWGLKVAEQLI 295
DB 241 YPOLVCLAVNLVFLVSIIRQ-LVTEKHQKQVDEVDFVFFVMTFLOFIFYMGWLKVEVLM 299
QY 296 NPFGEDEDDFETNWIIVDRNLQVSLAVDEMHDLPMEPDYMNKPEPQPPYTAASAOQR 355
DB 300 NPFGEDEDDFETNALIDRNITWGLKXVONT-MKTPELLKQDFDEVLVLLYSESSQIS 358
QY 356 RASFNGSTFNISLNK--EMEPQNOEDE 382
DB 359 NYHYHGSTSEVHLEQKCSVRMIPHSQE 387


```
Db 195 VPEEISIH2DDRMKYLPLTFMLGFFVTTVPERWRSALNVNPFIESVALSVAVLLPKGRE 254
QY 121 SRLRLRTLIRYANLGNVLILRSVSTAVYKRFPSQAHLVQAGFMTPAEHKOLEKLSLPHNM 180
Db 255 DRLTRRAIRYVVLHOLIVFRDISMRVRFPFTLYKVVVDAGFWRQBELDVLESVNOESSQ 314
QY 181 -FWPFWWFWANLSMKAWLGRIRDPILLQSLNEMNTLTQCCHLYAYDWISIPLYVTQV 239
Db 315 TWVFPINWANSALVAHQOKLIDQPTAFNNVIPAIFRVMETLIKFDIPAIPAIYPOV 374
QY 240 VTWAVYSFPLTCLVGRQFL-NPAKAYPGHELDLVVPVETFLQFFVYVGMKVAEQLINPF 298
Db 375 VFLAVRVYPAICLVSRQFLISDMKSKT--QMDWPFVIMTVLEFIFVIGMKVAEVLNLF 432
QY 299 GEDDDDFETNIVDRNLQVLSLAVDEMHDQLPRMEPDMYN----- 339
Db 433 GEDDDDFEVNSIIDNNISRGMAIVDTTHGVDPDLVDVSDPNLPAYSNSQIPNLTG 492
QY 340 -----KPEPQPPYTAASQF-----RRASFNGST----- 363
Db 493 SAAKVELAAPTDEVKIVRVNPDAPPTSRGSMFTRRNAYSLSRNKISIASNNLESPOE 552
QY 364 --FNIS-----LNKEEMEPQNOEDEDAHAGIIGRFLGLQSHDHPPRANSRTKLLWPK 416
Db 553 RKFNLSMPAGMLNKSTQDPDRPTMETVSEHE-----PSHFYRGDRVHSSDSVIPSF 603
QY 417 RES 419
Db 604 RRS 606

RESULT 15
B88710
protein C43G2.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Mar-2002
C:Accession: B88710
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: B88710
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAB09111.1; PID:g1572760; GSPDB:GN00022; CESP:C43G2.
C:Genetics:
A:Gene: C43G2.4
A:Map position: 4
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK849.5

Query Match 19.9%; Score 622; DB 2: Length 420;
Best Local Similarity 31.7%; Pred. No. 1.2e-40;
Matches 151; Conservative 82; Mismatches 140; Indels 104; Gaps 16;

QY 1 MTITYTSQVANARLGSFSLRLCWRGSIYKLLYGELIFLLCYIIRF----- 48
Db 1 MTISYS-----GNPFLLLRWKGSIRWSVWRELFLFLFYFIRESPHFFNYTDPT 52
QY 49 -----IYRLALTEEQOLMEKUTLYCDSVIQIPISFVLGTYVTLVVTWNNQYENLP 101
Db 53 DSKGYRKIFKVMCNFHE-----YTKMPLTLFLGTYVSVNVSRRWQFETLR 100
QY 102 WPDRLMSLVSGFVRGDKQSRLLRRTLIRYANLGNVLILRSVSTAVYKRFPSQAHLVQAG 161
Db 101 WPEDFLSILCLLLPSK--ESRPAHQIARYUNLTSALAWRDVSTKIRLPSPLENIIDAG 158
QY 162 FMTPAEHKOLEKLSLPHNMFWV--PWVWFANL-SMKAWLGGRIKDPPI-LQSL----- 211
Db 159 LLTEKEVEKLQDINVSKAVILMLPVYTFKNLCETEPSGIRWLTLPLHWVQQLIDAEITA 218
```

```
QY 212 -----NEMNTLRTQCCHLYAYDWISIPLYVTQVTVAVYSFPLTCLVGRQPLN 259
Db 219 GRGSVNYVSVAITNELKAYRISFRLLYCHDWVCVPLVITQVAALATYSYFFCLFGRODLN 278
QY 260 PAKAYPGHELDLVVPVETFLQFFVYVGMKVAEQLINPFGEDEDDDDFETNIVDRNLQVSL 319
Db 279 HDDFY---SLDAFFPLFTVVQFLFFVGVFKVQGLMRPFGLDLDDDDFELSVIDRNIWTSF 335
QY 320 LAVDEMH-QDLPRMSPDMYNNKPEPQPPYTAASQFRRASFNGSTFNISLNKEEMFQPN 378
Db 336 TIVDSLQDDDDPKFBEDVFWKHNEQ-----QQKHQSMFL-PRVPTSLKMGRIIDLSRN 387
QY 379 QDEDEDAHAGIIGRFLGLQSHDHPPRANSRTKLLWPKRESLLHEGLPKNHKAACKON 435
Db 388 -----AH-----KHPPKLTLYLEM-----KNQD-----PEYKRNKKN 415
```

Search completed: July 10, 2003, 12:27:09
Job time : 20.4626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:24:19 ; Search time 22.3425 Seconds
(without alignments)
3048.258 Million cell updates/sec

Title: US-09-622-964-3
Perfect score: 3120
Sequence: 1 MTITVTSQVANRLGSFRL.....TLKDHMPYALENDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 segs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCI NEW PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	44.8	261	9	US-09-746-783-4
2	661	21.2	314	10	US-09-768-826-54
3	539	17.3	251	10	US-09-768-826-35
4	238	7.6	99	9	US-09-764-872-429
5	111	3.6	361	9	US-10-114-893-88
6	104.5	3.3	786	10	US-09-815-242-10103
7	103.5	3.3	454	9	US-10-156-761-12378
8	99.5	3.2	516	9	US-09-866-050A-662
9	98	3.1	410	9	US-10-149-819-5
10	97.5	3.1	713	10	US-09-801-368-408
11	97.5	3.1	1384	9	US-10-226-315-2
12	97	3.1	846	9	US-10-005-691-2
13	97	3.1	1798	9	US-10-176-847-96
14	97	3.1	2061	9	US-10-036-036-181
15	96.5	3.1	465	9	US-10-029-180-44
16	94.5	3.0	1736	10	US-09-919-497-98
17	94	3.0	1247	9	US-09-908-193-2
18	93.5	3.0	624	9	US-10-156-239-24
19	93.5	3.0	624	9	US-10-199-485-24

Sequence 24, Appl
Sequence 342, App
Sequence 341, App
Sequence 2, Appl
Sequence 19, Appl
Sequence 22, Appl
Sequence 19, Appl
Sequence 12, Appl
Sequence 54, Appl
Sequence 110, App
Sequence 21, Appl
Sequence 21, Appl
Sequence 220, App
Sequence 2, Appl
Sequence 4, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 27, App
Sequence 43, App
Sequence 20, Appl
Sequence 303, App
Sequence 11, Appl
Sequence 4, Appl
Sequence 3135, Ap
Sequence 8135, Ap

624 10 US-09-795-693-24
894 9 US-09-884-465A-342
900 9 US-09-884-465A-341
1885 10 US-09-920-346-2
659 9 US-10-211-962-19
980 9 US-09-908-193-22
980 9 US-09-908-193-19
1252 9 US-10-118-513A-12
2243 9 US-10-176-847-54
2828 9 US-10-177-293-110
2828 9 US-09-905-129-21
2828 10 US-09-991-630-21
626 10 US-09-765-272-220
2016 9 US-09-896-994-2
932 10 US-09-754-997A-4
1162 9 US-10-226-579-2
1252 9 US-09-908-193-21
1252 10 US-09-754-997A-2
898 9 US-10-043-487-277
1162 9 US-10-079-625-43
1253 9 US-09-908-193-20
1569 9 US-10-108-605-303
855 9 US-10-245-175-11
1177 9 US-10-193-692-4
2192 9 US-10-128-714-3135
2209 9 US-10-128-714-8135

ALIGNMENTS

RESULT 1

US-09-746-783-4
; Sequence 4, Application US/09746783
; Publication No. US20030044935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Racie, Lisa A.
; Treacy, Maurice
; Spaulding, Vikki
; Agostino, Michael J.
; Howes, Steven H.
; Fectel, Kim

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/746,783

FILING DATE: 21-Dec-2000

CLASSIFICATION: unknown

ATTORNEY/AGENT INFORMATION:

NAME: Milasincic, Debra J.

REGISTRATION NUMBER: 46,931

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 261 amino acids

```
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-746-783-4

Query Match
Best Local Similarity 44.8%; Score 1397; DB 9; Length 261;
Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 325 MHQDLPRMPEPMYWKPEPPYTAASAFRRASFGSTFNISLNKEEMEFQPNQDEED 384
Db 1 MHQDLPRMPEPMYWKPEPPYTAASAFRRASFGSTFNISLNKEEMEFQPNQDEED 60

QY 385 AHAGIIGRFLGQSHDHPHPRANSRTKLLWPKESSLHHEGLPKVHKAAKQVGRQEDNKA 444
Db 61 AHAGIIGRFLGQSHDHPHPRANSRTKLLWPKESSLHHEGLPKVHKAAKQVGRQEDNKA 120

QY 445 WKLKAVDAFKSGPLYQRPYISAPOTPLSPTPMFFPLEPSAPSKLHSVTGIDTKKSLKT 504
Db 121 WKLKAVDAFKSAPLYQRPYISAPOTPLSPTPMFFPLEPSAPSKLHSVTGIDTKKSLKT 180

QY 505 VSSGAKSPFLLSESQALMEHPVQVRRKTVFNLTMPEIPENHLKEPLEQSPNTIH 564
Db 181 VSSGAKSPFLLSESQALMEHPVQVRRKTVFNLTMPEIPENHLKEPLEQSPNTIH 240

QY 565 TTLKDHMDPYWALENDRDEAHS 585
Db 241 TTLKDHMDPYWALENDRDEAHS 261

RESULT 2
US-09-768-826-54
; Sequence 54, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; FILE REFERENCE: P512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 54
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-54

Query Match
Best Local Similarity 21.2%; Score 661; DB 10; Length 314;
Matches 136; Conservative 31; Mismatches 53; Indels 32; Gaps 6;

QY 172 EKLSPHNMFVWPVWFWANLSMKVLGGRIRDPILLOSLNEMWTLRTQCGHLXAYDWIS 231
Db 13 ESKSDFNKYVPCVWFNTLAAQARRDGRIRDIALLCLLESLNKYRAKCSMLFHYDWIS 72

QY 232 IPLVYTVQVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLWVPVF 276
Db 73 IPLVYTVQVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE-----LDLWVPVF 132

QY 277 TFLQFFFTVGMKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLLAVDHMDHODLPRMPEPM 336
Db 133 TLLQFFFTVGMKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLLAVDHMDHODLPRMPEPM 192

QY 337 YWNKEPEPPYTT-AASAFRRASFGSTFNISLNKEEMEFQPNQDEEDAHAG----- 388
Db 193 YWDEDPQPPYTVATAESLRSPFLGSTFNLRMSDD-----PEQSLQVEASPGSRPAPA 247
```

```
QY 389 -----IIGREFLGL 396
Db 248 AQTPLUGREFLGV 259
```

```
RESULT 3
US-09-768-826-35
; Sequence 35, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; FILE REFERENCE: P512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-35
```

```
Query Match
Best Local Similarity 17.3%; Score 539; DB 10; Length 251;
Matches 111; Conservative 26; Mismatches 31; Indels 32; Gaps 6;

QY 224 LYAYDMISIPLVYTVQVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE----- 268
Db 2 LFHYDMISIPLVYTVQVTVAVYSFFLTCLVGRQFLNP-----AKAY-----PGHE----- 61

QY 269 LDLVVPVFTFLQFFYVGMKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLLAVDHMDHOD 328
Db 62 PMYVPLTLLQFFYVGMKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLLAVDHMDHOD 121

QY 329 LPRMPEPMYWKPEPPYTT-AASAFRRASFGSTFNISLNKEEMEFQPNQDEEDAHAG 387
Db 122 LPPAEKQYWDQDQPPYTVATAESLRSPFLGSTFNLRMSDD-----PEQSLQVEASPG 176

QY 388 G-----IIGREFLGL 396
Db 177 GSGRPAPAAQTPLUGREFLGV 196
```

```
RESULT 4
US-09-764-872-429
; Sequence 429, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 429
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-429
```

```
Query Match
Best Local Similarity 7.6%; Score 238; DB 9; Length 99;
Matches 46; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 MTITYTSQVANARLGSFSLLCWRGSIYKLLYGEFLIFLCYVYIFRFLALTEQQL 60
```

Db 13 MTVSYTKVAERFGGSGLLLEWRGSIYKLLYKEFLFGALYAVLSITVRLLLTQORY 72
Qy 61 MFPEKLTLYCDSYQILPISPLVG 83
Db 73 VYAQVARYCNRSADLPLSFVLG 95

RESULT 5

US-10-114-893-88
; Sequence 88, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, S.
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 88
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (273)
US-10-114-893-88

Query Match 3.6%; Score 111; DB 9; Length 361;
Best Local Similarity 24.4%; Pred. No. 0.15;
Matches 64; Conservative 35; Mismatches 97; Indels 66; Gaps 10;
Qy 341 PEOPPYTAA-----SAQFRASFMGSTFNI-----SLNKEEMEFQPNQDEDEDAHAGIG 391
Db 107 PETQPLMLSTADKSDSSSPERASASQSTKLRPSLQKPSI----- 148
Qy 392 RFLGLQSHDHPHPRANSTKLLMPKR--BSLLHEGLPKQHKAAKQNVRGQED-----NKA 444
Db 149 -----PNSAGKLTSPYKPSKSKSKSKSLKSHLEKAHKGSLEKCLKLQYA 197
Qy 445 WKLKAVDAFKSGPLQYQRYGYSACOTPLSPTEPMPFPPLPSAPSLSVGTIDTKDKSLKT 504
Db 198 CKLASSDKPVAPPOLFKPLYSHPQNEISPSKPGPQQLAKPKPHN-----PKRS 248
Qy 505 VSSGAKSKSFLLSSSDGALMEHPVSQVRRKTVENPLTDMPEIPENHLKEPLEQSPNTI 564
Db 249 VSLGRA--rALLSNSLA-----ETCQPKKKXKLVAKYRPLV--NDISEAKEKNTQNLH 298
Qy 565 TT--LKHMDPYWALENRDEAH 584
Db 299 VSSKVKSSRSRFRKLDGRNNAY 320

RESULT 6

US-09-815-242-10103
; Sequence 10103, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohleen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10103
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10103

Query Match 3.3%; Score 104.5; DB 10; Length 786;
Best Local Similarity 21.1%; Pred. No. 1.9;
Matches 69; Conservative 51; Mismatches 134; Indels 73; Gaps 16;
Qy 31 LLYGEPFLILCYVIIRIYRL-----ALTEBQQLMFEKLT-----YC 69
Db 231 MIIGAFTIDLLALLTLFVGQVLSNLAGSRTTAFQOSFLNAPALIEFFKAVRLRIFC 290
Qy 70 DSYIQLIPISFVLGPFYVTVVTRWNNQYENLPWDRMLSVS-GFVEKDEQSRLRLRTL 128
Db 291 PNVAELRE-----FTIODESARYWSR--RLSW---LSSLYGLIVAVPIISQVN--- 336
Qy 129 IRYANLGNVLIRSVST-AVYKRFPSAQHLVQAGFMTPAEHKQLEKLSLPHNMFVWVW 187
Db 337 VQIGALANVIIMLCWTVWALYIFRNKKEITQ-HLLNPAEH-SLAFFSLFIRAFALVWHM 394
Qy 188 PANL-----SMKANLGGRIIRDPILL-----QSLNEMNTLRTQCG 222
Db 395 LASAYFIVLPFPLSFDGNSLKFMMGATVRSIAITIGIAAFVSGMFSRWLAKTITLSPTQ 454
Qy 223 HLY-----AYDWTISIPLVTVQVTVVAVYSFPLCLVGR-QFLNPAKAYPGHE-LDLVVP 274
Db 455 RNYPELQKRLNGWLSAALKATRIILTVCVAVMILLISANGLFDVWNLQNGAGQKTVDIR 514
Qy 275 VFTPLQPFVYVWMLKVAEQLINPEGED 301
Db 515 IALIL-EFSAVGVTVLASLIENRLASD 540

RESULT 7

US-10-156-761-12378
; Sequence 12378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HOKIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

```
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12378
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-12378

Query Match
Best Local Similarity 3.3%; Score 103.5; DB 9; Length 454;
Matches 72; Conservative 38; Mismatches 101; Indels 97; Gaps 18;

QY 9 VANARIGSFRLLCW-RGSIYKLLYGEFLIFLLCYIIRFYRLALTEQQQLMFEKLT 67
Db 139 LAVASVIGIVLAGSSGTYPLLG-----LRSCQMIS--YEIAM-----GRAFASVFL 188
QY 68 YCDS-----YIQLIPISFVLGYVTLVTRWNOYENLPW--PDRMLSLVS 111
Db 189 YSGSMSTSTVEQOHRWYIVLLPVSVFI--YIVTMV-----GETNRAPFDPMPSEGLVG 242
QY 112 GFVEGKDEQSRLRLRLIRYANLGNVILRSVSTAVYKRPSPSAHLVQAGFMTPAEHKQL 171
Db 243 GF-----NTYSSIKFAMFMAEYVNMVTSVSTTLF-----LGGWRAP----- 282
QY 172 EKLSLPHNFW-----PWVFA-----NLSMKAWLGG-----RIRDPILLO----- 208
Db 283 -----WPISFEGANHGWMFMLFVVKVQLLFFFLWGLTLPVRYDQMLKLGKVLIP 338
QY 209 -----SLNEMNTLTQCHLYADWISIPLYVTVVTVVAVYSPFLTCLVGRQFINPAK 262
Db 339 VSVVWMLVATVTKLRNE-----NYDFADIAL-----YVGGVLLVLLLSFVADMPREKSK 389
QY 263 --AYFCHE 268
Db 390 EAAAPAE 397

RESULT 8
US-09-866-050A-662
; Sequence 662, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 662
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Rat
; US-09-866-050A-662

Query Match
Best Local Similarity 3.2%; Score 99.5; DB 9; Length 516;
Matches 54; Conservative 30; Mismatches 105; Indels 57; Gaps 8;

; APPLICANT: PRMEPDYMNKPEPQPPYTAASQAQFRRAFMGTFNLSNKEEMFQPNQDEDEDAHAGI 389
; 101 PALQSTSKQPTPKFTALVTRGTHKPTSEGLE---SVGPVAPDFEPTSTOHLATSKV 157
; 390 IGRFLGLQSH--DHHPPRANSRTKLLWPKRESILHEGLPKNHKAAKQNVRGQEDNKAWKL 447
; 158 TQSLLTQSSPVASVSTTPKLPVPIAQLTLEPVQTSRORRRATKQGSRT--- 214
; 448 KAVDAFKSGPLYQRPQYYSAPQTPMFFPLEPSAPSKLHSVTGIDTKOKSLKTVSS 507
; 215 -----APVGPKSYSYPAEP-----EPOSSASQSS 238
; 508 GAKKSPFLLSESDGALMEHPEVSQVRKTVENLTDMPENHNLKEPLEQSPNTNHTTL 567
; 239 GA-----SEADSPHQKRP-RQVTKQTVVVKBEDPGEI---QVKE--EPOETAISTPG 285
; 568 KDHMDP 573
; 286 KRKRDP 291

RESULT 9
US-10-149-819-5
; Sequence 5, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dying Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PP-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 1689337CD1
; US-10-149-819-5

Query Match
Best Local Similarity 3.1%; Score 98; DB 9; Length 410;
Matches 72; Conservative 25; Mismatches 133; Indels 76; Gaps 15;

QY 330 PRMEPDYMNKPE-----PQPPYTAASQAQFRRAFMGTFNLSNKEEMFQPNQDEDEDAHAGI 370
Db 98 PEKEERRVWTWPMVAVALKPVLOOSREARDELPGAPPVLCSSSDLSLLGLGPF-----Q 152
QY 371 EEMFQPNQDEDEDAHAGIIGRF--LGLQSHDHHPPRANSRTKLLWPKRESILHEGLPKN 428
Db 153 SOHSFQP-LERPDLTSSTAGAFSALGAFPHDRAERPFEED--PGPDGEGLLKQGLPPA 210
QY 429 HKAQKQNVRCQ-----EDNKAWKLKAVDAF--KSGPLYQRPQYYSAPQTPMFFPP 480
Db 211 QLEGLKNFLHQLLETVPQNNEN---PSVDLLPPKSGPL--TVPSWEAPQVPRIPPVHK 266
QY 481 LEPSA-----PSKLHSVTGIDTKOKSLKTVSSGAKSKFELLSESD----- 520
```


Db 267 KVLAMASSLFRVPEPPSSHS-QSGSPSGSPERGGLTTPRQLMEVSQLRLYQARGW 325
QY 521 GAL-----MEHEVSVQRKTVFNLTD--MPEIPENHLKEPLEQSPNTIHTTL 567
Db 326 GALPAEDLLLYLKLHSGRTDGRDNVPRNTDSRLGEI--PSQAVPRRLATAP 383
QY 568 KDHMDP 573
Db 384 KTEKPP 389

RESULT 10

US-09-801-368-408
; Sequence 408, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 408
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-408

Query Match 3.1%; Score 97.5; DB 10; Length 713;
Best Local Similarity 23.1%; Pred. No. 7;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASAFRRASFMSGTFTNISLNKEEMFQPNQDEE-----DAHAGI 389
Db 18 AIRQEFQVQSQANTYRLQ-NOKDYDFKNNQQLAEMQIIRNTVYELETHRRKMKDAYEE 76
QY 390 IGRF-LGLOSHDHHPPRANSRTKLLWPKRSLHGLPKNHKAQKONVREGQEDNKAWLK 448
Db 77 IGHKLGLQERD-----QIASLTQV---QORQQQQQQVQOHHQQQQQLA 120
QY 449 AVDAKSGPLYQR-PGYVSAPQTPISPT-----PMFFPLEPSAPSKLHSGVTGIDTKDKSLK 503
Db 121 AASA--SVEVAQPPATTSATPAANTTTGSPSAFFVQASRP-----NLVGSQLPITLIP 174
QY 504 TVSSCAKSKFELLSSDGNALMEH---PEVSQVRKTVFN-----LTDMPFI--PENH 551
Db 175 VVSSNAQQQLPQQQLQQOQLOQQPPQVSVAPLNTAINGSPTSKEITTLPSVKAPEST 234
QY 552 LKEPLEQSPNTIHTT 566
Db 235 LK---ETEPENNTS 246

RESULT 11

US-10-226-315-2
; Sequence 2, Application US/10226315

Publication No. US20030120051A1
; GENERAL INFORMATION:
; APPLICANT: Poles, Elior
; TITLE OF INVENTION: CASPR/190, A FUNCTIONAL LIGAND FOR RPTP-BETA AND THE
; TITLE OF INVENTION: AXONAL CELL RECOGNITION MOLECULE CONTACTIN
; FILE REFERENCE: 7683-111
; CURRENT APPLICATION NUMBER: US/10/226,315
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 08/826,134
; PRIOR FILING DATE: 1997-03-26
; PRIOR APPLICATION NUMBER: 60/014,199
; PRIOR FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-226-315-2

Query Match 3.1%; Score 97.5; DB 9; Length 1384;
Best Local Similarity 20.5%; Pred. No. 19;
Matches 95; Conservative 50; Mismatches 180; Indels 139; Gaps 21;
QY 203 DPILQSLNEMNTLTTCQCHLYAYDWISIPLYVTQV-----TVAVYSFF---LTLV 253
Db 725 DPALYCNCADQPOWRTDKGLLTFVDHLPV---TQVIGDTNRTSEAQFFLRPLRCYG 780
QY 254 GROFLNPAKAYPG-----HELDLVVPVFTFLQFFP-----YVG 286
Db 781 DRNSWNTISFHTGAALRFPPIRANHSLD-----VSFVFTSAPSGVLENMGPGYCQ 832
QY 287 W-----LKVAEQLINPF---GEDD-----DDEFTN---WIVDR---NLQVSLAVD 323
Db 833 WRPPYRVVELNTRDVFADVGNGDENTVHSDDEFNDEHVLVRAEINVKQALRVD 892
QY 324 EMHQDLPRMPEPMYNNKPEPPYTAASAQFRRASFMSGTFTNISLNKEEMFQPNQDEE 383
Db 893 HRPVTLRPMPLQTYIMWEYDQPLY-VGSALKRPPVGCRLAMRLNGVTL---NLEGRA 947
QY 384 DAHAGIIGRFLGLOSHDH-----HPPRANSRTKLLWPKRSLHGLPKNHKAQKONVREGQ 439
Db 948 NASEGTSNCTGCHCAHPLRPFCHGRCVRYSYTYTCDCLTAFDGPYCNH---DIGGF 1002
QY 440 HONKAWK-----LKAVDAFKSG--PLYQRPVY-----YSAPQT 470
Db 1003 FPGTWMRYNLOSALRSAREFSHMSLRPVPQYEPGYIPGYDTPGYVPGYHGGVRLPDY 1062
QY 471 PLSPPTMFFPLEPSAPSKLHSGVTG-----IDTKDKSLKTVSGCAKSKFELLSSDGA 522
Db 1063 PRPGRPV-----PGYRGVYNTVGTSEVSFSSTSSAPAVLLYVSFVRDYMVAVLKDDGT 1117
QY 523 LMEHEVSVQRKTVFNLTDMPFIENHLKEPLEQSPNTIHTT 566
Db 1118 LQ-----LRYQLGTSPPYVYQLTTRPVTGQPHSINIT 1149

RESULT 12

US-10-005-691-2
; Sequence 2, Application US/10005691
; Publication No. US20020183250A1
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: MICHALOVICH, DAVID
; TITLE OF INVENTION: NOVEL USE
; FILE REFERENCE: GH-30003-D1
; CURRENT APPLICATION NUMBER: US/10/005,691
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 09/107,847
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: EP 97304996.8
; PRIOR FILING DATE: 1997-07-08
; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 846
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-005-691-2

Query Match      3.1%; Score 97; DB 9; Length 846;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 45; Conservative 28; Mismatches 57; Indels 46; Gaps 10;

QY 416 KRESLLHGLPKNHAKQNVGOEDNKAWL---KAVDAFKSGPL-----YORPGYYS 466
Db 386 RRELIGIESLPET--AADKQDSGSDNRINTVLSKEALERFDHSPTSPASRSRKSST 443
QY 467 APOTPLSPPTMFFLEPSAPSKLHSVTGIDTKOKSLKTVSSGAKKSPILLSSESDGALMEH 536
Db 444 AVSDP-SSTPTKPTDSTTPRQH---LPAHEKWQRRSPFSQSIN--SQSVGSLSLQ 496
QY 527 PEVSQVRKTVFENLTMPEIPE-----NHLKEPLEOS-----PTNIH 564
Db 497 PVMSQA-----TNLP-IPQMSQFSAQLGAMQHLKDQLEQRTMRMEANIH 542

RESULT 13
US-10-176-847-96
; Sequence 96, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petteer Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 1798
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-96

Query Match      3.1%; Score 97; DB 9; Length 1798;
Best Local Similarity 19.5%; Pred. No. 32;
Matches 105; Conservative 73; Mismatches 188; Indels 172; Gaps 29;

QY 102 WPDRLMSLVSGFVEGKDEQSRLLRRLTRYANLGNVLILRSVSTAVYKFPSPSAQHLVQAG 161
Db 494 WLDKLMQLTE---EPQNSMPDII-----IMWIRGEKRLAYARIPAHQVL--- 534
QY 162 FMTPAEHLKQLEKLSLPHNMF-----VPMWVFANLSMKAWLGGRIIDPILLOSL 210
Db 535 YSTSGENASGKYCGKTQTIFLYKPOEKNGPKVP-----VELRVNIWLG-----LSAV 582
QY 211 LNMNMTLRQCCHLYAYDWISIPLVYTVVAVYSFFLTCLVGRQFLNPAKAYPGHELD 270
Db 583 EKKFNSFAEGTFTVFAEMYNQALMFGKWT-----SGLVGR-----HKFS 623
QY 271 LVVPVFTFLQFFFY--VGLWKVAEQLINPFGEDDDDDFETNWIVD--RNLOVSLA--- 321
Db 624 DVTGKIKLKRFFLPKGM-----EWEGEIVDPERSLLTTEADAGHTEF 667
QY 322 VDEMHDQLPRMEPDMYNNKPEPPYTAASAOFRRASFMGSTFNISLNKEEMEFOPNOED 381
Db 668 TDEVYQNESRY-PGGDW-KP-AEDTYTDANG-----DKAASPELTCPPGWEW 712
QY 382 BEDAHAGIIGRL-----GLQSHDHPHPRANSRTKLLWPKRESLLH-----EGLPKNHK 430
Db 713 EDDAWSYDINRAVDKMGWEYGITIPPDHKPKS-----WVAAEQWYTHRRRLVRKRK 765
QY 431 -----AAQNVRGOE---DNKAWLKAVDAPK-----SGPLYQRPGYYSAPQTPLSPTP 476
```

```
Db 766 KDLTQTASSTARAMELQOEGWEYASLIGWKFWKQSRSDTFRRRW----- 813
QY 477 MFFPLEPSAPSKLHSVTGIDTKOKSL--KTVSSGAKKSPILLSSESDGALM--EHPVSOV 532
Db 814 ----RRKMAPSETHGAAAIKLEGALGADTTEDGDEKSLKQKHSATTVFGANTPIVSCN 869
QY 533 RRKTVFENL-----TDMPEIPENHLKEP-----LEQSPTN--IHTTLKDHMDPYW 575
Db 870 FDRVYIHLRCVYQARNLLALDKDSFDPYAHICFLHRSKTTETIHTSL-----NPTW 923

RESULT 14
US-10-060-036-181
; Sequence 181, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yuguu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 2061
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-181

Query Match      3.1%; Score 97; DB 9; Length 2061;
Best Local Similarity 19.5%; Pred. No. 39;
Matches 105; Conservative 73; Mismatches 188; Indels 172; Gaps 29;

QY 102 WPDRLMSLVSGFVEGKDEQSRLLRRLTRYANLGNVLILRSVSTAVYKFPSPSAQHLVQAG 161
Db 757 WLDKLMQLTE---EPQNSMPDII-----IMWIRGEKRLAYARIPAHQVL--- 797
QY 162 FMTPAEHLKQLEKLSLPHNMF-----VPMWVFANLSMKAWLGGRIIDPILLOSL 210
Db 798 YSTSGENASGKYCGKTQTIFLYKPOEKNGPKVP-----VELRVNIWLG-----LSAV 845
QY 211 LNMNMTLRQCCHLYAYDWISIPLVYTVVAVYSFFLTCLVGRQFLNPAKAYPGHELD 270
Db 846 EKKFNSFAEGTFTVFAEMYNQALMFGKWT-----SGLVGR-----HKFS 886
QY 271 LVVPVFTFLQFFFY--VGLWKVAEQLINPFGEDDDDDFETNWIVD--RNLOVSLA--- 321
Db 887 DVTGKIKLKRFFLPKGM-----EWEGEIVDPERSLLTTEADAGHTEF 930
QY 322 VDEMHDQLPRMEPDMYNNKPEPPYTAASAOFRRASFMGSTFNISLNKEEMEFOPNOED 381
Db 931 TDEVYQNESRY-PGGDW-KP-AEDTYTDANG-----DKAASPELTCPPGWEW 975
QY 382 BEDAHAGIIGRL-----GLQSHDHPHPRANSRTKLLWPKRESLLH-----EGLPKNHK 430
Db 976 EDDAWSYDINRAVDKMGWEYGITIPPDHKPKS-----WVAAEQWYTHRRRLVRKRK 1028
QY 431 -----AAQNVRGOE---DNKAWLKAVDAPK-----SGPLYQRPGYYSAPQTPLSPTP 476
Db 1029 KDLTQTASSTARAMELQOEGWEYASLIGWKFWKQSRSDTFRRRW----- 1076
QY 477 MFFPLEPSAPSKLHSVTGIDTKOKSL--KTVSSGAKKSPILLSSESDGALM--EHPVSOV 532
Db 1077 ----RRKMAPSETHGAAAIKLEGALGADTTEDGDEKSLKQKHSATTVFGANTPIVSCN 1132
QY 533 RRKTVFENL-----TDMPEIPENHLKEP-----LEQSPTN--IHTTLKDHMDPYW 575
```

```

RESULT 15
US-10-029-180-44
; Sequence 44, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffry C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-44

```

Query Match	3.1%;	Score 96.5;	DB 9;	Length 465;
Best Local Similarity	21.2%;	Pred. No. 4.5;		
Matches	91;	Conservative 56;	Mismatches 144;	Indels 139; Gaps 25;
QY	154	AQHLVQAQFMTPAEHKOLEKLSLPHNNFWVPVWFANLSMKWMLGG--RIRQPIILQSL	210	
DB	89	ASHAAQMGFARGAQMQQ-QQLHQNH-----GRLAETKAGGVKTRIRD-VWKHNL	137	
QY	211	LNEMWTLRTQCGHLIYADWHSIPLVYQVTVVAVYSF-----FLTCLVG-----	254	
DB	138	AQEMAILRQLQ--VEKYPYISMDTEFGIVARPIGAFTNKADYHQTLCNCVLLKMIQL	194	
QY	255	--RQFLNPAPKAYPGCHELDL-----VVPVFIFLOEFFYVGMVKAQELINPGEDDDD	304	
DB	195	GITLFSSEGEVPPNATDANGQPLGNGLVPAPCTWQFN-----RPSUEDDM	241	
QY	305	FETNWTVDRLNQLVSLAIVDMHQDLPRMEP-----DMYWNKPEPOPPYT	348	
DB	242	YAO-----ESTAMLAKAGIDPFAMHDKNGIDPFEGALLISSGLVLDDVHW-----	287	
QY	349	AASAQFRRASFMSGSTFNISLNKEEMEFQPNQDEEDAHAGIIIGRFGL-----QSHD-----	400	
DB	288	---VSPHSGYDFGLMKIMLCK-----PUPNEEEFHF-----RLNIFFPSLYDIKYL	333	
QY	401	HHPPRANSTRKL-----LWPKRESLLHEGLPKNHKAQKONVRQGDNKAWLKAV-----	450	
DB	334	KHAGRNQAASGLQDIADGLGVKRVGIAHQ-----GSDSLVTGE---IYWKMRQLVFNG	384	
QY	451	---DAFKSGPLY,---ORQG--YYSAP-CTP-LSPTFMPFPL-EPASPKLSHSVTGIDTK	498	
DB	385	NIDEAKYSQIIGLNGQMPALLYSMQPHOTPNLNGATIYSATGTPTPNAVHSVTGSHTP	444	
QY	499	DKSLKTVSSG	508	
DB	445	QHALTPGATG	454	

Search completed: July 10, 2003, 12:28:57
Job time : 24.3425 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	100	3.2	2183	3	US-08-746-111-5	Sequence 5, Appli
2	99.5	3.2	713	1	US-08-190-802A-63	Sequence 63, Appl
3	99.5	3.2	713	4	US-08-477-346-63	Sequence 63, Appl
4	99.5	3.2	713	4	US-08-473-089-63	Sequence 63, Appl
5	99.5	3.2	713	4	US-08-487-072A-63	Sequence 63, Appl
6	99.5	3.2	713	4	US-09-108-857-3	Sequence 3, Appli
7	99	3.2	342	3	US-08-785-928-1	Sequence 1, Appli
8	99	3.2	342	3	US-08-728-603-17	Sequence 17, Appl
9	98.5	3.2	681	2	US-08-272-255-6	Sequence 6, Appli
10	98.5	3.2	681	2	US-08-272-255-6	Sequence 9, Appli
11	98.5	3.2	681	5	PCT-US95-08565-6	Sequence 6, Appli
12	98.5	3.2	681	5	PCT-US95-08565-9	Sequence 9, Appli
13	98	3.1	192	4	US-09-134-001C-4329	Sequence 9, Appli
14	97.5	3.1	1384	4	US-08-526-131-2	Sequence 4329, Ap
15	97	3.1	846	3	US-08-985-291-55	Sequence 2, Appli
16	97	3.1	846	3	US-09-107-847-2	Sequence 55, Appl
17	97	3.1	846	4	US-09-496-672-55	Sequence 2, Appli
18	95.5	3.1	681	3	US-08-964-268-5	Sequence 55, Appl
19	95.5	3.1	870	4	US-09-134-001C-4959	Sequence 5, Appli
20	95	3.0	1139	4	US-09-046-158A-22	Sequence 4959, Ap
21	94.5	3.0	655	2	US-08-272-255-18	Sequence 22, Appl
22	94.5	3.0	655	5	PCT-US95-08565-18	Sequence 18, Appl
23	93.5	3.0	765	2	US-09-395-689-1	Sequence 18, Appl
24	93.5	3.0	765	2	US-08-663-112-2	Sequence 1, Appli
25	92.5	3.0	659	4	US-09-562-737-19	Sequence 2, Appli
26	92	2.9	626	4	US-08-961-083-220	Sequence 19, Appl
27	92	2.9	1132	2	US-08-567-508C-2	Sequence 220, App

Db 908 PSELIPKQKTSKFLNRWRVASEKGSVEIIAANGEDTDVXLTNSPQONITV----- 962
Qy 323 DEMHODLRMEPMYWN---KPEPPOPYTAASQFRASFMGSTFNISLNKEEMEFQPNQ 379
Db 963 -----PRGESTHTNTRKPSDLPTPSGVGHK-----SPHVRQEEENSFGKQK 1006
Qy 380 -----EDEEADAHAGIIGR-FLGLOSHDHP---PRANSRTKLLWPKESSLHEGL 425
Db 1007 LFIRTRKKKX 1066
Qy 426 PKNHKAQKQNRGQDNKAWKLVKAVDAFKSGPLYQRPYYSAPQTPLESTPMFFPLEPSA 485
Db 1067 NOTSPMSMSTRSLPDYQYKNDTEOMSSLDLYQ-----SVPAEHSPT---FPAQD--- 1116
Qy 486 PSKLHVSU-----GID-----TKOKSLKT-----VS 506
Db 1117 PDQTHSTDPYSRSPPELSQLDYDLSDHFDYDDIGLTSFFPDQSKSFSFSDDDQAIP 1176
Qy 507 SGAKSKFELLSESGALMEHPEVSQV-----RRKTVEENLTDMPKPEIENHLK 553
Db 1177 SSDLSLFTISPDLDTII-YPOLDQLLSPEDNQKTSPPDLGGVPLSPDNDOK 1228

RESULT 2

US-08-190-802A-63
; Sequence 63, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P. O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 01-FEB-1994
; APPLICATION NUMBER: US/08/190,802A
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUP1, Fig. 46
US-08-190-802A-63

Query Match 3.2%; Score 99.5; DB 1; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;

Qy 349 AASAFRRASFMGSTFNISLNKEEMEFQPNQDEE-----DAHAGI 389
Db 18 AIRQFLVQSQANTYRIQ-NQKDYDFKXNQQLAEMQIQNTVYELTHRKMKADEAE 76
Qy 390 IGRF-LGLOSHDHPHPRANSRTKLLWPKESSLHEGLPKNHKAQKQNRGQDNKAWKLV 448
Db 77 IKHLKLGLEQRDH-----QIASLTVQ--QQQQQQQQQQVQOHLQQQQQOLA 120
Qy 449 AYDAKSGPLYQR-PGYYSAPOTPLSPT-----PMPEPLEPSAPSKLHVSVDITKOKSLK 503
Db 121 AASA--SVEVAQPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSOLPTTLP 174
Qy 504 TVSSGAKSKFELLSESGALMEH---PEVSQVRKKTVEFN-----LTDMPKPEI--PENH 551
Db 175 VVSSNAQQQLPQOQLQQOQLQQOQLQQOQLQQOQLPPQVSVAPLSNTAINGSPSKETTLPLSVKAPEST 234
Qy 552 LKEPLEQSPTNIHTT 566
Db 235 LK---ETEPENNTS 246

RESULT 3

US-08-477-346-63
; Sequence 63, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUP1, Fig. 46
US-08-477-346-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;

QY 349 AASQFRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEANTYRLQ-NOKDYDFKMNQQLAEQOIRNTVYELTHRKMKAAYEAE 76
QY 390 IGRF-LGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWKLK 448
DB 77 IKHLKLGLEQRDH-----QIASLTQV--QOQQOQQOQVOQHLQOQQOQLA 120
QY 449 AVDAFKSGPLYQOR-PCYYSAPOTPLSPT-----PMFPFLPSAPSAPSKLHVSVTGIDTKDKSLK 503
DB 121 AASA--SVPAQOPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSQPLTTTLP 174
QY 504 TVSSGAKGSELLSBSGALMEH---PEVSQVRKKTVEFN-----LTDMPPI--PENH 551
DB 175 VVSSNAQQLPOOQLOOQOQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPEST 234
QY 552 LKEPLEQSPNTNIHTT 566
DB 235 LK---ETEPENNTS 246

RESULT 4

US-08-473-089-63
; Sequence 63, Application US/08473089
; Patent No. 6342368
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,089
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUPI, Fig. 46
US-08-473-089-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASQFRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEANTYRLQ-NOKDYDFKMNQQLAEQOIRNTVYELTHRKMKAAYEAE 76

QY 390 IGRF-LGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWKLK 448
DB 77 IKHLKLGLEQRDH-----QIASLTQV--QOQQOQQOQVOQHLQOQQOQLA 120
QY 449 AVDAFKSGPLYQOR-PCYYSAPOTPLSPT-----PMFPFLPSAPSAPSKLHVSVTGIDTKDKSLK 503
DB 121 AASA--SVPAQOPPATTSATATPAANTTTGSPSAFPVQASRP-----NLVGSQPLTTTLP 174
QY 504 TVSSGAKGSELLSBSGALMEH---PEVSQVRKKTVEFN-----LTDMPPI--PENH 551
DB 175 VVSSNAQQLPOOQLOOQOQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPEST 234
QY 552 LKEPLEQSPNTNIHTT 566
DB 235 LK---ETEPENNTS 246

RESULT 5

US-08-487-072A-63
; Sequence 63, Application US/08487072A
; Patent No. 6423684
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,072A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUPI, Fig. 46
US-08-487-072A-63

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
QY 349 AASQFRASFMGSTFNISLNKEEMFQPNQDEE-----DAHAGI 389
DB 18 AIRQEFLOVSQEANTYRLQ-NOKDYDFKMNQQLAEQOIRNTVYELTHRKMKAAYEAE 76
QY 390 IGRF-LGLQSHDHPHPRANSRTKLLWPKRSLHLEGLPKNHKAQKONVRQEDNKAWKLK 448

Db 77 IKHLKLGLEQRDH-----QIASLTQV--QQQQQQQQQQVQOHLQQQQQOLA 120
Qy 449 AVDAFKSGPLYQR-PGYYSAPQTPLSPT-----PMFFPLEPSAPSKLHVTGIDTKDKSLK 503
Db 121 AASA--SVPAQOPPATTSATATPAANTTGTSPSAPFVQASRP-----NLVGSQLPPTTLP 174
Qy 504 TVSSGAKKSFELLESQALMEH---PEVSQVRKRTVEFN-----LTDMPET--PENH 551
Db 175 VVSSNAQQQLPQOOLQOQQOQQPPQVSVAPLSNTAINGSPTSKEKTTTLPSSVKAPEST 234
Qy 552 LKEPLEQSPTNIHTT 566
Db 235 LK---ETEPENNNTS 246

RESULT 6

US-09-108-857-3
; Sequence 3, Application US/09108857A
; Patent No. 643137
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, Alexander D.
; APPLICANT: BRAUN, Burkhard
; TITLE OF INVENTION: TUP1 SEQUENCES FROM CANDIDA ALBICANS AND METHODS FOR
; SCREENING AGENTS FOR INHIBITING VIRULENCE IN CANDIDA
; TITLE OF INVENTION: ALBICANS USING TUP1
; FILE REFERENCE: 22002-20006.00
; CURRENT APPLICATION NUMBER: US/09/108,857A
; CURRENT FILING DATE: 1998-07-01
; EARLIER APPLICATION NUMBER: 60/051,552
; EARLIER FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-108-857-3

Query Match 3.2%; Score 99.5; DB 4; Length 713;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 59; Conservative 38; Mismatches 95; Indels 63; Gaps 13;
Qy 349 AASAQFRASFQMGSTFNISLNKEEMFQPNODEE-----DAHAGI 389
Db 18 ATRQEFLOVQSEANTRLQ-NQKDYDFKMQQAEQQQINTVYELTHRKKMKDAYEAE 76
Qy 390 IGRF-IGLQSHDHPPRANSRTKLLWPQRESLLHGLPKNHKAQKQVRQEDNKAWKLK 448
Db 77 IKHLKLGLEQRDH-----QIASLTQV--QQQQQQQQQQVQOHLQQQQQOLA 120
Qy 449 AVDAFKSGPLYQR-PGYYSAPQTPLSPT-----PMFFPLEPSAPSKLHVTGIDTKDKSLK 503
Db 121 AASA--SVPAQOPPATTSATATPAANTTGTSPSAPFVQASRP-----NLVGSQLPPTTLP 174
Qy 504 TVSSGAKKSFELLESQALMEH---PEVSQVRKRTVEFN-----LTDMPET--PENH 551
Db 175 VVSSNAQQQLPQOOLQOQQOQQPPQVSVAPLSNTAINGSPTSKEKTTTLPSSVKAPEST 234
Qy 552 LKEPLEQSPTNIHTT 566
Db 235 LK---ETEPENNNTS 246

RESULT 7

US-08-785-928-1
; Sequence 1, Application US/08785928
; Patent No. 6087115
; GENERAL INFORMATION:
; APPLICANT: Gershengorn, Marvin C.
; APPLICANT: Arvanitakis, Leandros
; APPLICANT: Geras-Raaka, Elizabeth
; APPLICANT: Cesarman, Ethel
; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED

; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
; ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,928
; FILING DATE: 22-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1320
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-785-928-1

Query Match 3.2%; Score 99; DB 3; Length 342;
Best Local Similarity 25.0%; Pred. No. 0.15;
Matches 70; Conservative 41; Mismatches 109; Indels 60; Gaps 15;
Qy 3 IYTSQVANARIGSPSRLLL--CWGSIYKL-LYGEFLIFLLCYIIIRLALTEBQQ 59
Db 68 VYIFCKRSRAGADIDLLGICLSICLSIAEVLMLFPNIIISTGLCRL-----E 121
Qy 60 LMFEKILYCDYSYIQLIPISFVLGPFYTVLVTRWNOYENLPWDRMLSLVSGFVEGKD- 118
Db 122 IFYYLYVYLDIF-SVVCVSLV-RYLLVAYSTRSPKQSLGWLTSAALLIALVLSGA 179
Qy 119 --EQSRL--LRRTLIRYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFWTPAEH 168
Db 180 CRHSRVVDVPSKQAMCYENAGNMTADWRLHVRTVS-----VTAGFLLP--- 223
Qy 169 KQLEKLSLPHNMFVPMVWTFANLSKAWLGRIIRDPILL-----QSLNEMNTL--- 217
Db 224 --LALLILFYALTWCV-VRRTKLQARRKRGVIVAVLLFFVFCPPHYHVLNLLDTLRRR 280
Qy 218 --RTQCGLHYADWISIPLYVT---QVTVVAVYSFFLTCL 252
Db 281 WIRDSC---YTRGLINVLAVTSLQALYSVVPLIYSL 317

RESULT 8

US-08-728-603-17
; Sequence 17, Application US/08728603
; Patent No. 6093806
; GENERAL INFORMATION:
; APPLICANT: Cesarman, Ethel
; APPLICANT: Knowles, Daniel M.
; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
; WITH HERPESVIRUS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; STREET: Clinton Square, P.O. Box 1051

APPLICANT: Lin, Chentao
TITLE OF INVENTION: Blue Light Photoreceptors and Methods of


```

/ TITLE OF INVENTION: Using the Same
/ NUMBER OF SEQUENCES: 22
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
/ STREET: One Liberty Place, 46th floor
/ CITY: Philadelphia
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/272,255
/ FILING DATE: 08-JUL-1994
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Leary Ph.D., Kathryn
/ REGISTRATION NUMBER: 36,317
/ REFERENCE/DOCKET NUMBER: UPN-1795
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 681 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-272-255-9

Query Match
Best Local Similarity 3.2%; Score 98.5; DB 2; Length 681;
Matches 93; Conservative 69; Mismatches 136; Indels 193; Gaps 27;

QY 123 LLRLTIRIYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLS 175
DB 267 LVRIQVAVANEGNEAGESVNLFLKSGIGREYSRYSFNH-----PYSH---ERPL 315
QY 176 LPHNFWVFWVWFWANLSKAWLGGRIIDPILLOSLLENWNTLRTOCGHLYAYDWTISIPLV 235
DB 316 LGHLKFF-PMWVDENY-FKAWRQGRGTGYP-LVDAGMRE-----LWATGWLH---- 358
QY 236 YQVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOPFF-----Y 284
DB 359 --DRIRVVVSFFVKV-----LQLPWVGWKYFWDTLDDLESDA 397
QY 285 VGWLKVA-----EQINPFGE-----DDDDFETNWIVDRLNQVSLVAVDEMHQDLPL 330
DB 398 LGWQVITGLPDSREFDRIDNPQEGYKDPNGEVVRRWLP-----ELSLPTDWHHP-- 451
QY 331 RNEPDWYKPKPPPYTAASQAFRRASFGWSTFNISLNKEEMFQPNQDEEDAHAGII 390
DB 452 ---WNAPE--SVLQAAGIE-----LGSNYPLPI----- 474
QY 391 GRFLGLQSHDHHPPRANSRTKLLWPKRSLHLEGLPKNHKAQKQNVGRQEDNKAKLKV 450
DB 475 ---VGLD-----EAKARLHEAL-----SOMQLEAA 497
QY 451 D--AFKSGPLYQRPVYSAPQPLSFTMFPF-----LEPSAFSKLH-----SVT 493
DB 498 SRAAIENG---SEEGIGLSAE--VBEAPTEFFPRDITMEETETRLNPNRRYEDQMVPSIT 552
QY 494 GI-----DTKDKSLKTVSGAKKSPEL-----LSESDGALME--HPEVSQVRKTVFNLTD 543
DB 553 SSLIPEDEESSLNURNSVGSRAEVPVRNMTVNOAQORRAEPASNQVTAMIPFNIRI 612
QY 544 MPEIPENHLKE 554
DB 613 VAESTEDSTAE 623

```

```

RESULT 11
PCT-US95-08565-6
/ Sequence 6, Application PC/TUS9508565
/ GENERAL INFORMATION:
/ APPLICANT: Cashmore, Anthony R.
/ APPLICANT: Ahmad, Margaret
/ APPLICANT: Lin, Chentao
/ TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
/ TITLE OF INVENTION: Using the Same
/ NUMBER OF SEQUENCES: 22
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
/ STREET: One Liberty Place, 46th floor
/ CITY: Philadelphia
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/08565
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/272,255
/ FILING DATE: 08-JUL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Leary Ph.D., Kathryn
/ REGISTRATION NUMBER: 36,317
/ REFERENCE/DOCKET NUMBER: UPN-1795
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 681 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ PCT-US95-08565-6

Query Match
Best Local Similarity 3.2%; Score 98.5; DB 5; Length 681;
Matches 93; Conservative 69; Mismatches 136; Indels 193; Gaps 27;

QY 123 LLRLTIRIYANLGNV-----LILRSVSTAVYKRFPSAQHLVQAGFMTPAEHKQLEKLS 175
DB 267 LVRIQVAVANEGNEAGESVNLFLKSGIGREYSRYSFNH-----PYSH---ERPL 315
QY 176 LPHNFWVFWVWFWANLSKAWLGGRIIDPILLOSLLENWNTLRTOCGHLYAYDWTISIPLV 235
DB 316 LGHLKFF-PMWVDENY-FKAWRQGRGTGYP-LVDAGMRE-----LWATGWLH---- 358
QY 236 YQVVTVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOPFF-----Y 284
DB 359 --DRIRVVVSFFVKV-----LQLPWVGWKYFWDTLDDLESDA 397
QY 285 VGWLKVA-----EQINPFGE-----DDDDFETNWIVDRLNQVSLVAVDEMHQDLPL 330
DB 398 LGWQVITGLPDSREFDRIDNPQEGYKDPNGEVVRRWLP-----ELSLPTDWHHP-- 451
QY 331 RNEPDWYKPKPPPYTAASQAFRRASFGWSTFNISLNKEEMFQPNQDEEDAHAGII 390
DB 452 ---WNAPE--SVLQAAGIE-----LGSNYPLPI----- 474
QY 391 GRFLGLQSHDHHPPRANSRTKLLWPKRSLHLEGLPKNHKAQKQNVGRQEDNKAKLKV 450
DB 475 ---VGLD-----EAKARLHEAL-----SOMQLEAA 497

```

QY 451 D--AFKSGPLYORPGYISAPOTPLSPTMFPF-----LEPSAPSKLH-----SVT 493
 DB 498 SRAAIENG---SEGLGDSAE--VEEAPIEFPRDITMEETETRLNPNRRYEDQWVPSIT 552
 QY 494 GI-----DTKDKSLKTVSSGAKKSPFEL---LSESDGALME--HPEYSQVRRKTVFENLTD 543
 DB 553 SSLIRPEDEESSLNLRNSVGDSRAEVPNRMVNTNQAOORRAEPASNOVTAMIPFNIRI 612
 QY 544 MPEIPENHLKE 554
 DB 613 VAESTEDSTAE 623

RESULT 12
 PCT-US95-08565-9
 ; Sequence 9, Application PC/TUS9508565
 ; GENERAL INFORMATION:
 ; APPLICANT: Cashmore, Anthony R.
 ; APPLICANT: Ahmad, Margaret
 ; APPLICANT: Lin, Chentao
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 ; TITLE OF INVENTION: Using the Same
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
 ; STREET: One Liberty Place, 46th floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/08565
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/272,255
 ; FILING DATE: 08-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary Ph.D., Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: UPN-1795
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 568-3100
 ; TELEFAX: (215) 568-3439
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 681 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 PCT-US95-08565-9

Query Match 3.2%; Score 98.5; DB 5; Length 681;
 Best Local Similarity 18.9%; Pred. No. 0.5; Indels 193; Gaps 27;
 Matches 93; Conservative 69; Mismatches 136; Indels 193; Gaps 27;

QY 123 LLRRTLIRYANLGNV-----LILRSYSTAVYKRPFSQAHLVQAGFTWPAEHKQLEKLS 175
 DB 267 LVRIKQVAMANEAGNEAGESVNLFLKSLGLREYSRYSFNH-----PVSH---ERPL 315
 QY 176 LPHNMFWVFWVWFAANLKNKAWLGRIRDPIILQSLNEMNLTQCCHLYAYDWISIPLV 235
 DB 316 LGHLKFF-PWAVDENY-FKAWROGRGTGYP-LVDAGMRE-----LWATGWLH----- 358
 QY 236 YTVQVTVAVYSFFFTCLVGRQFLNPAKAYPGHELDLVVVFPTPLQFPF-----Y 284

DB 359 --DRIRVVVSSPFVKV-----LQLPWRWGMKYFWDTLDDADLESDA 397
 QY 285 VGWLKVA-----EOLINPGE-----DDDDPETNWIVDRNLQVSLLAVDMEHQDLP 330
 DB 398 LQWQVITGLPSRFRDRIDNPFQEGYKPDNGEYVRRWLP-----ELSLRFTDWIHP-- 451
 QY 331 RMEPDYMNKPPQPPYTAASQAFRRASFMGSTFNISLNKEEMEFQPNQOEDEDAAGII 390
 DB 452 -----WNAPE--SVLQAAGIE-----LGSNYPPI----- 474
 QY 391 GRFLGLQSHDHPHPRANSRTKLLWPKRESLLHGGIPONHKAQKONVGRQEDNKAWLKAV 450
 DB 475 ---VGLD-----BAKARLEAL-----SOMQLEAA 497
 QY 451 D--AFKSGPLYORPGYISAPOTPLSPTMFPF-----LEPSAPSKLH-----SVT 493
 DB 498 SRAAIENG---SEGLGDSAE--VEEAPIEFPRDITMEETETRLNPNRRYEDQWVPSIT 552
 QY 494 GI-----DTKDKSLKTVSSGAKKSPFEL---LSESDGALME--HPEYSQVRRKTVFENLTD 543
 DB 553 SSLIRPEDEESSLNLRNSVGDSRAEVPNRMVNTNQAOORRAEPASNOVTAMIPFNIRI 612
 QY 544 MPEIPENHLKE 554
 DB 613 VAESTEDSTAE 623

RESULT 13
 US-09-134-001C-4329
 ; Sequence 4329, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4329
 ; TYPE: PRT
 ; LENGTH: 192
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4329

Query Match 3.1%; Score 98; DB 4; Length 192;
 Best Local Similarity 24.8%; Pred. No. 0.076;
 Matches 51; Conservative 24; Mismatches 71; Indels 60; Gaps 11;

QY 287 WLKVAEQI-INPFGEDDDPETNWIVDRNLQVSL-----LAYDEMHQDLPRMEPDYMNKPP 341
 DB 15 YIEEAKDLVINSIGETMDLYGINSVGNLYGIMVPEGSMTLDMEKQQLQMSKPSNSTGVK 74
 QY 342 EPQPPYTAASQAQFERAS-----FMGS-----TP--NISLNKEEMEFQPNQOEDEDAAGII 391
 DB 75 KLQ-EYDIVVQKQFTRGSRKOHFAEKDFFTFFRNFTTKCKWEREIDTNNEATEDA----- 127
 QY 392 RFLGLQSHDHPHPRANSRTKLLWPKRESLLHGGIPKN---HKAAKQ--NVRGQEDNK--- 443
 DB 128 -----ENIHPLEKKNIDIEEVKQKAINVKAQLDHSKIY 161
 QY 444 -ANKLKAVDAFKGGLYQRCGYISAP 468
 DB 162 YKMLAQLSEALESGEIF---NYFFIP 184

RESULT 14
 US-08-826-134-2
 ; Sequence 2, Application US/08826134A

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:19:13 ; Search time 36.1203 Seconds
(without alignments)
2158.111 Million cell updates/sec

Title: US-09-622-964-3

Perfect score: 3120

Sequence: 1 MTITYTQVANALGFSFRL.....TLKDHMPYALENDEAHS 585

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.2

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3120	100.0	585	20	Human CG1CE long f
2	1899	60.9	551	20	Mouse CG1CE protei
3	1535	49.2	435	20	Human CG1CE short
4	1397	44.8	261	20	Human adult retina
5	1009	32.3	721	22	Drosophila melanog
6	886	28.4	535	22	Drosophila melanog
7	824.5	26.4	368	22	Novel human diagno
8	742	23.8	327	22	Drosophila melanog
9	694.5	22.3	731	22	Drosophila melanog
10	661	21.2	314	22	Human secreted pro

11	549	17.6	188	23	AB889288	Human polypeptide
12	539	17.3	251	22	AA870067	Human secreted pro
13	539	17.3	251	23	AB865506	Human albumin fusi
14	523	16.8	101	22	ABB14768	Human nervous syst
15	324.5	10.4	701	22	AB808608	Novel human diagno
16	238	7.6	99	22	AAU22652	Novel human colon
17	238	7.6	99	22	AAU22658	Human digestive sy
18	205.5	6.6	377	22	AB806411	Novel human diagno
19	174	5.6	30	20	AAU29956	Human CG1CE partia
20	126	4.0	47	22	AB808607	Novel human diagno
21	118	3.8	30	20	AAU29971	C. elegans CG1CE r
22	117	3.8	30	20	AAU29968	C. elegans CG1CE r
23	115	3.7	30	20	AAU29969	C. elegans CG1CE r
24	112	3.6	1256	22	AB860532	Drosophila melanog
25	111.5	3.6	637	22	AAU23162	Novel human enzyme
26	111.5	3.6	637	22	AAU23701	Novel human enzyme
27	111.5	3.6	832	22	AAO13516	Human polypeptide
28	111.5	3.6	849	19	AA882397	Human UBP protein
29	109	3.5	283	21	AA842835	Human ORFX ORF2599
30	109	3.5	1051	21	AA842838	Arabidopsis thalia
31	109	3.5	1057	21	AA842837	Arabidopsis thalia
32	107	3.4	384	21	AA826452	Drosophila melanog
33	107	3.4	384	22	AB866539	Drosophila melanog
34	107	3.4	389	21	AB820926	Drosophila odorant
35	107	3.4	722	22	AB865012	Drosophila melanog
36	106	3.4	581	18	AAW22103	Murine leptin rece
37	105	3.4	518	22	AAW39378	Human polypeptide
38	105	3.4	531	22	AAW41164	Human polypeptide
39	105	3.4	1816	21	AAV95440	Caenorhabditis ele
40	104.5	3.3	785	23	AB891027	Herbicidally activ
41	104.5	3.3	786	22	AAU34510	E. coli cellular p
42	103.5	3.3	356	22	AB893521	Human protein sequ
43	103.5	3.3	356	23	AB897330	Novel human protei
44	103	3.3	30	20	AAU29970	C. elegans CG1CE r
45	103	3.3	854	21	AA842329	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAU29953
ID AAU29953 standard; Protein; 585 AA.
XX
AC AAU29953;
XX
XX
DT 22-NOV-1999 (first entry)
XX
DE Human CG1CE long form protein sequence.
XX
KW CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;
KW BMD; age-related macular dystrophy.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 121
FT /note= "encoded by GGC"
XX
PN WO9943695-A1.
XX
PD 02-SEP-1999.
XX
PF 22-FEB-1999; 99WO-US03790.
XX
PR 25-FEB-1998; 98US-0075941.
PR 18-DEC-1998; 98US-0112926.
XX
PA (MERI) MERCK & CO INC.
XX (UYUP-) UNIV UPPSALA.
XX
PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX

DR WPI; 1999-540560/45.
 DR N-PSDB; AA221227.
 PT Human and mouse polynucleotides encoding CG1CE polypeptides -
 XX Claim 7; Fig 3; 67pp; English.
 XX The present sequence represents the human CG1CE protein. When the CG1CE
 CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
 CC Polynucleotides encoding CG1CE are useful for diagnosing whether a
 CC patient carries a mutation in the CG1CE gene. Normal and mutated
 CC CG1CE proteins are useful for identifying activators and/or inhibitors
 CC of these proteins, in order to treat BMD. The CG1CE gene offers a
 CC simpler and cheaper method of diagnosing BMD without the need for the
 CC presence of the patient. The gene may also be useful to discovering
 CC the genetic cause of age-related macular dystrophy.
 XX Sequence 585 AA;
 SQ Query Match 100.0%; Score 3120; DB 20; Length 585;
 Best Local Similarity 100.0%; Pred. No. 5e-296;
 Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEQQQL 60
 DB 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEQQQL 60
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 DB 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 QY 121 SLLLRRTLIRYANLGNVLIRSVSTAVYKRFPSAQLVQAGFMTPAEHKQLEKLSLPHNM 180
 DB 121 SLLLRRTLIRYANLGNVLIRSVSTAVYKRFPSAQLVQAGFMTPAEHKQLEKLSLPHNM 180
 QY 181 FWVPWFWFANLGMKAWLGGRIIDPILLOSLLNEMTLRTCCGHLAYDWMISIPLVTTQVV 240
 DB 181 FWVPWFWFANLGMKAWLGGRIIDPILLOSLLNEMTLRTCCGHLAYDWMISIPLVTTQVV 240
 QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 DB 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 QY 301 DDDDFETNWIIDRNLOVSLAVDEMHDLPMEPDWYNNKPEPQPPYTAASQFRASFM 360
 DB 301 DDDDFETNWIIDRNLOVSLAVDEMHDLPMEPDWYNNKPEPQPPYTAASQFRASFM 360
 QY 361 GSTFNISLNKEEMEPQNPQE----DEDAHAGIIGRFLGLQSHDHPHPPRANSTRKLLWPK 416
 DB 361 GSTFNISLNKEEMEPQNPQE----DEDAHAGIIGRFLGLQSHDHPHPPRANSTRKLLWPK 416
 QY 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWKLKAVDAFKSGPLVQRPYGYSAPOTPISPTP 476
 DB 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWKLKAVDAFKSGPLVQRPYGYSAPOTPISPTP 476

XX CG1CE; Best's macular dystrophy; mutation; diagnosis; detection;
 KW BMD; age-related macular dystrophy.
 XX Mus sp.
 XX WO9943695-A1.
 PN 02-SEP-1999.
 XX 22-FEB-1999; 99WO-US03790.
 XX 25-FEB-1998; 98US-0075941.
 PR 18-DEC-1998; 98US-0112926.
 XX (MERI) MERCK & CO INC.
 PA (UYUP-) UNIV UPPSALA.
 XX Petrukhin K, Caskey CT, Metzker M, Wadelius C;
 PI WPI; 1999-540560/45.
 DR N-PSDB; AA221229.
 XX Human and mouse polynucleotides encoding CG1CE polypeptides -
 PT Claim 7; Fig 8; 67pp; English.
 XX The present sequence represents the mouse CG1CE protein. When the CG1CE
 CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
 CC Polynucleotides encoding CG1CE are useful for diagnosing whether a
 CC patient carries a mutation in the CG1CE gene. Normal and mutated
 CC CG1CE proteins are useful for identifying activators and/or inhibitors
 CC of these proteins, in order to treat BMD. The CG1CE gene offers a
 CC simpler and cheaper method of diagnosing BMD without the need for the
 CC presence of the patient. The gene may also be useful to discovering
 CC the genetic cause of age-related macular dystrophy.
 XX Sequence 551 AA;
 SQ Query Match 60.9%; Score 1899; DB 20; Length 551;
 Best Local Similarity 64.3%; Pred. No. 1.3e-176;
 Matches 378; Conservative 59; Mismatches 101; Indels 50; Gaps 10;
 QY 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEQQQL 60
 DB 1 MTITTSQVANARLGSFLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEQQQL 60
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 DB 61 MFEKLTLYCDSYIQLIPISFVLGFFVTVVTRWNNQYENLPWDRMLSVSGFVEGKDEQ 120
 QY 121 SLLLRRTLIRYANLGNVLIRSVSTAVYKRFPSAQLVQAGFMTPAEHKQLEKLSLPHNM 180
 DB 121 SLLLRRTLIRYANLGNVLIRSVSTAVYKRFPSAQLVQAGFMTPAEHKQLEKLSLPHNM 180
 QY 181 FWVPWFWFANLGMKAWLGGRIIDPILLOSLLNEMTLRTCCGHLAYDWMISIPLVTTQVV 240
 DB 181 FWVPWFWFANLGMKAWLGGRIIDPILLOSLLNEMTLRTCCGHLAYDWMISIPLVTTQVV 240
 QY 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 DB 241 TVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVTFLOFFPYVGNLKVAEOLINPFG 300
 QY 301 DDDDFETNWIIDRNLOVSLAVDEMHDLPMEPDWYNNKPEPQPPYTAASQFRASFM 360
 DB 301 DDDDFETNWIIDRNLOVSLAVDEMHDLPMEPDWYNNKPEPQPPYTAASQFRASFM 360
 QY 361 GSTFNISLNKEEMEPQNPQE----DEDAHAGIIGRFLGLQSHDHPHPPRANSTRKLLWPK 416
 DB 361 GSTFNISLNKEEMEPQNPQE----DEDAHAGIIGRFLGLQSHDHPHPPRANSTRKLLWPK 416
 QY 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWKLKAVDAFKSGPLVQRPYGYSAPOTPISPTP 476
 DB 417 RESLLHEGLPKNHKAQKQNVRCQEDNKAWKLKAVDAFKSGPLVQRPYGYSAPOTPISPTP 476

RESULT 2
 ID AAY29955
 AC AAY29955 standard; Protein; 551 AA.
 XX AAY29955;
 XX AAY29955;
 DT 22-NOV-1999 (first entry)
 DE Mouse CG1CE protein sequence.

Db 421 NPLL--BQCKD-----ANQNKQD--VWFKGLDFLKVPRFKRSGHCGQAPSS--- 468

Qy 477 MFPPLPSASKLSHVTGIDTKDLSLKTSSGAKKSPFLLSESGALMEHPEVSVRRKT 536

Db 469 --HPTQSAFSS--SDTG-----DGPSTDYQEIChMKKKT 499

Qy 537 VEFNLTDMPETPENHLKE-PLSQSPNTHTTLKDHMPYWALENRDEA 583

Db 500 VEFNL-NIPESPTEHLQORLDQMSNTIQALMKHAESY---PYRDEA 543

RESULT 3

AA29954

ID AAY29954 standard; Protein; 435 AA.

XX AC AAY29954;

XX DT 22-NOV-1999 (first entry)

XX DE Human CGICE short form protein sequence.

XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;

XX KW BMD; age-related macular dystrophy.

XX OS Homo sapiens.

XX PN WO9943695-A1.

XX PD 02-SEP-1999.

XX PF 22-FEB-1999; 99WO-US03790.

XX PR 25-FEB-1998; 98US-0075941.

XX PR 18-DEC-1998; 98US-0112926.

XX PA (MERI) MERCK & CO INC.

XX PA (UYUP-) UNIV UPPSALA.

XX PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;

XX DR WPI; 1999-540560/45.

XX DR N-PSDB; AA221228.

XX Human and mouse polynucleotides encoding CGICE polypeptides

PS Claim 7; Fig 5; 67pp; English.

CC The present sequence represents the human CGICE protein. When the CGICE gene is mutated it is responsible for Best's macular dystrophy (BMD). Polynucleotides encoding CGICE are useful for diagnosing whether a patient carries a mutation in the CGICE gene. Normal and mutated CGICE proteins are useful for identifying activators and/or inhibitors of these proteins, in order to treat BMD. The CGICE gene offers a simpler and cheaper method of diagnosing BMD without the need for the presence of the patient. The gene may also be useful to discovering the genetic cause of age-related macular dystrophy.

SQ Sequence 435 AA;

Query Match 49.2%; Score 1535; DB 20; Length 435;

Best Local Similarity 99.7%; Pred. No. 3.7e-141;

Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTITTSQVANARLGSFRLLLCWRGSIYKLYGFEFLPLLCYIIRFYRLALTEEQQL 60

Db 1 MTITTSQVANARLGSFRLLLCWRGSIYKLYGFEFLPLLCYIIRFYRLALTEEQQL 60

Qy 61 MFEKLTLYCDSYIQLIPISFVLGPFVTVVVTWNNQYENLPDPRLMSLVGFVGKDEQ 120

Db 61 MFEKLTLYCDSYIQLIPISFVLGPFVTVVVTWNNQYENLPDPRLMSLVGFVGKDEQ 120

Qy 121 SRLLRRTLIYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Db 121 GRLLRRTLIYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHNM 180

Qy 181 FWVPWVWFANLSWKAMLGGRIRDPILLOSILNENMTLRTQCGHLYAYDWISIPLVYTVQV 240

Db 181 FWVPWVWFANLSWKAMLGGRIRDPILLOSILNENMTLRTQCGHLYAYDWISIPLVYTVQV 240

Qy 241 TVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFFVVGWLKV 290

Db 241 TVAVYSPFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFFVVGWLKV 290

RESULT 4

AAW95345

ID AAW95345 standard; Protein; 261 AA.

XX AC AAW95345;

XX DT 26-APR-1999 (first entry)

XX DE Human adult retina secreted protein bk112_15.

XX KW Secreted protein; human; retina; bk112_15.

XX OS Homo sapiens.

XX PN WO9856909-A2.

XX PD 17-DEC-1998.

XX PF 08-JUN-1998; 98WO-US11822.

XX PR 05-JUN-1998; 98US-0092722.

XX PR 11-JUN-1997; 97US-0873218.

XX PA (GEMY) GENETICS INST INC.

XX PI Agostino MJ, Fichtel K, Howes SH, Jacobs K, LaVallie ER;

XX PI McCoy JW, Racie LA, Spaulding V, Treacy M;

XX DR WPI; 1999-080899/07.

XX DR N-PSDB; AAV99722.

XX PT New polynucleotides encoding secreted human proteins - derived from human foetal brain, adult testes, foetal kidney, adult thyroid or adult retina cDNA libraries

PS Claim 13; Page 72-73; 113pp; English.

CC This is the amino acid sequence of bk112_15, a novel human secreted protein predicted from the nucleotide sequence of a human adult retina cDNA clone (see AAV99722). Database searches indicate some sequence similarity to known sequences. The invention provides cDNA clones (see AAV99721-33) from human adult thyroid, adult retina, adult testis, foetal kidney and foetal brain that encode novel secreted proteins (see AAW95344-53). The polynucleotides and proteins are predicted to have activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional, cytokine, cell proliferation/differentiation, immune stimulating (e.g. as vaccines) or immune suppressing, haematopoiesis regulating, tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic, thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour invasion suppressor, and tumour inhibition activities.

SQ Sequence 261 AA;

Query Match 44.8%; Score 1397; DB 20; Length 261;

Best Local Similarity 99.6%; Pred. No. 5.4e-128;

Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 325 MHQDLPRMEPDMYWNKPQPPYTAASAQFRASFMGSTFNISLNKEEMFQPNQDEED 384

Db 181 AARPPYTLAAADYCIPIFLSGSTVQMGSGDFPDEBWLWDYKGGHRRHSMIRRVKFLSA 240
Qy 397 QSHDHPHPRANSRTKL-----LWPKESLHGLPKN--HKAARKONVGOEDNKAWKLKA- 449
Db 241 HEHPSRRRSYRRRTSSSMFLPRDDLSHSGPTGCALKPKPPQGLTHLEDNPASQKEAP 300
Qy 450 --VDAFKSGPLYQRRG---YYSAPQTPLSPTPMFFPPLPSAPSKLSHVSVTGIDTKDKSLK 503
Db 301 RCTSAWESCPPSGRPARQALYRAXPHSPVKELP-----PSKCHGTGKGIDHSRST 351
Qy 504 TVSSGAKSKFELL 516
Db 352 AHIRGLPPXFRYL 364

RESULT 8
ABB63135
ID ABB63135 standard; Protein; 327 AA.
XX AC ABB63135;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 16197.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL07238.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 16197; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 327 AA;
Query Match 23.8%; Score 742; DB 22; Length 327;
Best Local Similarity 46.3%; Pred. NO. 9.1e-64;
Matches 142; Conservative 55; Mismatches 92; Indels 18; Gaps 4;

Qy 68 YCDSYLIQIPISFVLGFTYTLVTRWVNYENLPWDRMLSLVSGFVEGKDEOSRLRET 127
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 YCEKNGSLPLSFVLGFTYTLVTRWVNYENLPWDRMLSLVSGFVEGKDEOSRLRET 62

Qy 128 LIRYANLGNLIRSVSTAVYKRFPSAQLVQAGFMTPAEHKQLEK-----SLPHNMFV 182
Db 63 ILRYVCLCOVIVFTWISPRVKRFPYTYQIIEAGFLENEKKIIEAMDQAFPSYPKH--W 120
Qy 183 VPMWPFANLSKAMJGGRIIRDPILOSLNENWNTLRTQCGHLVAYDWISIPLYVTVVTV 242
Db 121 MBIVWAASTVWRARRENKIRDDYAVKTIIDELNLRGNCGLLYYDWISVPLVYTVVTV 180
Qy 243 AVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVGLKVAEQLINPGEED 302
Db 181 ATYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVGLKVAEQLINPGEED 240
Qy 303 DDFETNWIIDRNLTQVSLAVDEMHQDLPRMEPDMYWNK-----PEPQP-----PYTAAS 351
Db 241 DDFELNWIIDRNLTQVSLAVDEMHQDLPRMEPDMYWNK-----PEPQP-----PYTAAS 351
Qy 352 AQFRAS 358
Db 301 AYMESQS 307

RESULT 9
ABB64914
ID ABB64914 standard; Protein; 731 AA.
XX AC ABB64914;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 21534.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL09017.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 21534; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 731 AA;

Query Match 22.3%; Score 694.5; DB 22; Length 731;
Best Local Similarity 33.7%; Pred. No. 1.3e-58;
Matches 174; Conservative 80; Mismatches 195; Indels 67; Gaps 12;

QY 78 ISFVLGFFVTLVTWNNQYENLWPDPRLMSLVSGFVEG-----KDEQSRLRRLLIRYA 132
DB :
1 MSFVLGFFVNLVKRWKEQYRLLPDPDTIALFISAAIPNSNGGVNNETGRLMRRNIMRYM 60
DB :
QY 133 NLGNVLILRSYATVYKFPSPASHLVQAGFMTPABHKOLEKLS--LPNNMFVPVWVFAN 190
DB :
61 VLAVYTITLQRISLRVKRFFPTQHLDVADGLMHSEMKFEALNQKSPSKYWPLVWATN 120
QY 191 LSMKAWLGGRTRDPILOSLLNMNTLTQCCHLYAYDWISIPLYVTVVTVVAVYSFLT 250
DB :
121 IINRAKDGLIASDHIVQTILVELSDIRRLGGLIGYDTCVPLVYTQVTLVLVITYEIA 180
QY 251 CLVGROFLNPAPKAYPGHE-LDLVVPVFTFLQFFVVGHLKVAEQLINPFGEEDDDDFETNW 309
DB :
181 ALLGRQLPNVLDNRSGREDPDLPFLFTVLOFVFGVWLKVAEVLINPFGEEDDDDIELNW 240
QY 310 IVDRNLQVSLLAVDHMDLPRMEPMYWNKPPEQP-PYTAASAQFRASFMSGSTFNISL 368
DB :
241 LIDRHKAANYMIVDEMHEEHPELLRDQWECVVPKDLPTVASBHYRKDEBKGAEKYV 300
QY 369 NKEWEFQPNQDEDAHAGII-GRFLGLOSHHHPPRANSRTKLLWPKRESLLHLEGLPK 427
DB :
301 KKEDA-----MYANIMPGGCKMLSDDVYADYESVDTMPVERKKNNMLVRQLSR 349
QY 428 NHKAAQNVRQEDNKAKWLKXADAFKSGLP-YORPGYYSAPOTPLSTPMPFFLEP--- 483
DB :
350 MGSMRSQST-----AYSSGGMPFNRRNLNSVYSPESGLPLTILQQOOLQ 394
QY 484 -----SAPSKLHSVTGIDTKDSLKTVSSGAKKSFELLSSDGALMEHPVSQVRR 534
DB :
395 QAHQQOAGSQPSK-SPLYGFVHRKSLRAQROLIKQNSKLINGLVNVAKTRPRI----- 448
QY 535 KTFBNITDMPEIENHLKEPLEOS-----PTNIHTT 566
DB :
449 -----PTPEVAKDGNNTDNATSVLMAPQQLSTT 476

RESULT 10
AAB70085
ID AAB70085 standard; Protein; 314 AA.
AC AAB70085;
XX
XX 14-MAY-2001 (first entry)
XX Human secreted protein-related polypeptide #6.
XX Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KW immunostimulant; cytotoxic; cardiac; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nontropic; anticoagulant; vaccine;
KW antialzheimers; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
KW neurological; infection.

OS Homo sapiens.
XX WO200112776-A2.
XX
XX 22-FEB-2001.
XX
XX 15-AUG-2000; 2000WO-US22350.
XX
XX 16-AUG-1999; 99US-0148759.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
XX

Query Match 21.2%; Score 661; DB 22; Length 314;
Best Local Similarity 54.0%; Pred. No. 7.2e-56;
Matches 136; Conservative 31; Mismatches 53; Indels 32; Gaps 6;

QY 172 EKSLPHNFWPWVWFANLSMAWLGGRTRDPILOSLLNMNTLTQCCHLYAYDWIS 231
DB :
13 ESKLSDFNKYVCVWFTNLAAQAARRDGRIRDIALCLLELNKYRAKCSMLPHYDWIS 72
QY 232 IPLVYTOVVTVAVYSFLTCLVGROFLNP---AKAY----PGHE-----LDLVVPVP 276
DB :
73 IPLVYTOVVTVAVYSFALSGLVGROFVEPEGAAKPOKLLKPGEPAPALGDDPMYVPLT 132
QY 277 TFLQFFPYVGLKVAEQLINPFGEEDDDDFETNIVDRNLQVSLLAVDHMDLPRMEPDM 336
DB :
133 TLLQFFFYAGWKVAEQLINPFGEEDDDDFETNQLIDRNLTQVSLLSVDVMYQNLPPAEKQ 192
QY 337 YNNKPEPPPYT-AAQAQFRASFMSGSTFNISLKGEMEPQNOEDEDAHAG----- 388
DB :
193 YWDEDQPPPYTVATAAESLRPSFLGSTFNLRMSDD-----PQSLQVEASPGSGRPAPA 247
QY 389 ----IIGRFGL 396
DB :
248 AQTPLLIGRFLGV 259

RESULT 11
ABB89288
ID ABB89288 standard; Protein; 188 AA.
XX ABB89288;
AC ABB89288;
XX
XX 24-MAY-2002 (first entry)
XX Human polypeptide SEQ ID NO 1664.
XX
XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.
XX WO200190304-A2.
XX

```
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
PI N-PSDB; ABL89697.
XX
DR WPI; 2002-122018/16.
XX
DR N-PSDB; ABL89697.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment, and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 11; SEQ ID NO 1664; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 188 AA;
Query Match 17.6%; Score 549; DB 23; Length 188;
Best Local Similarity 97.2%; Pred. No. 3e-45;
Matches 105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 477 MFPPLEPSAPSKLHSGTIDTKDKSLKTVSSGAKKSFELLSESDGALMEHPEVSQVRKKT 536
DB 1 MFPPLEPSAPSKLHSGTIDTKDKSLKTVSSGAKKSFELLSESDGALMEHPEVSQVRKKT 60
QY 537 VFNLTDMPEIPENHLKEPLEQSPNTNHTTLKDHMDPYWALENRDEAH 584
DB 61 VFNLTDMPEIPENHLKEPLEQSPNTNHTTLKDHMDPYWALENRSLVH 108
RESULT 12
AAB70067
ID AAB70067 standard; Protein; 251 AA.
XX
AC AAB70067;
XX
DT 14-MAY-2001 (first entry)
XX
DE Human secreted protein #6.
XX
KW Human; secreted protein; immunomodulatory; anti-sclerotic;
KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
KW immune disorder; hyperproliferative; cardiovascular; angiogenic;
KW neurological; infection.
XX
OS Homo sapiens.
```

```
XX WO200112776-A2.
PN
XX
PD 22-FEB-2001.
XX
XX
PF 15-AUG-2000; 2000WO-US22350.
XX
XX
PR 16-AUG-1999; 99US-0148759.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
XX
XX
DR WPI; 2001-244245/25.
XX
DR N-PSDB; AAF76848.
XX
PT Nucleic acids encoding 18 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX
PS Claim 11; Page 362-363; 380pp; English.
XX
XX
CC The present sequence is one of 18 novel human secreted proteins. The
CC nucleic acids encoding the proteins and the proteins themselves may be
CC used in the prevention, diagnosis and treatment of diseases including
CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
CC and human immunodeficiency virus (HIV) infections), hyperproliferative
CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC neovascularisation and diabetic retinopathy), neurological disorders
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC infectious diseases and/or for promoting wound healing, regeneration
CC and/or chemotaxis. The nucleic acid molecules may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples. The polypeptides may also be
CC used as antigens in the production of antibodies and in assays to
CC identify modulators of protein expression and activity.
XX
XX
SQ Sequence 251 AA;
Query Match 17.3%; Score 539; DB 22; Length 251;
Best Local Similarity 55.5%; Pred. No. 4.4e-44;
Matches 111; Conservative 26; Mismatches 31; Indels 32; Gaps 6;
QY 224 LYAYDWISIPLVYTVQVTVAVYSFFLTCLVGRQFLNP----AKAY-----PGHE----- 268
DB 2 LFHYDWISIPLVYTVQVTVAVYSFFALSILVGRQFVEPEAGAKPKQLKKEGQEPAPALGD 61
QY 269 LDLVVPVFTFLQFFYVGMVKVAEQILNPFGBDDDDDFETNWIIVDRNLQVSLLAIDEMHQD 328
DB 62 PMYVPLATLLQFFYVGMVKVAEQIINPFGBDDDDDFETNQLIDRNLQVSLSDVEMYQN 121
QY 329 LPRMEPDWYMKPEQPYPYV--AASQFRASFWGSTFNISLNKEEFQPNQDESDAHA 387
DB 122 LPPAEKQYWDQDQPPYTVATAESLSPFLSGTFLNLRMSDD-----PEQSLQVEASP 176
QY 388 G-----IIGRFLGL 396
DB 177 GSGRPAPAAQTPLGLRFLGV 196
RESULT 13
ABG65506
ID ABG65506 standard; Protein; 251 AA.
XX
AC ABG65506;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human albumin fusion protein #2181.
XX
KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
```


PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0232402.
 PR 14-SEP-2000; 2000US-0232403.
 PR 14-SEP-2000; 2000US-0232404.
 PR 14-SEP-2000; 2000US-0232405.
 PR 14-SEP-2000; 2000US-0232406.
 PR 14-SEP-2000; 2000US-0232407.
 PR 14-SEP-2000; 2000US-0232408.
 PR 14-SEP-2000; 2000US-0232409.
 PR 14-SEP-2000; 2000US-0232410.
 PR 14-SEP-2000; 2000US-0232411.
 PR 14-SEP-2000; 2000US-0232412.
 PR 14-SEP-2000; 2000US-0232413.
 PR 14-SEP-2000; 2000US-0232414.
 PR 14-SEP-2000; 2000US-0232415.
 PR 14-SEP-2000; 2000US-0232416.
 PR 14-SEP-2000; 2000US-0232417.
 PR 14-SEP-2000; 2000US-0232418.
 PR 14-SEP-2000; 2000US-0232419.
 PR 14-SEP-2000; 2000US-0232420.
 PR 14-SEP-2000; 2000US-0232421.
 PR 14-SEP-2000; 2000US-0232422.
 PR 14-SEP-2000; 2000US-0232423.
 PR 14-SEP-2000; 2000US-0232424.
 PR 14-SEP-2000; 2000US-0232425.
 PR 14-SEP-2000; 2000US-0232426.
 PR 14-SEP-2000; 2000US-0232427.
 PR 14-SEP-2000; 2000US-0232428.
 PR 14-SEP-2000; 2000US-0232429.
 PR 14-SEP-2000; 2000US-0232430.
 PR 14-SEP-2000; 2000US-0232431.
 PR 14-SEP-2000; 2000US-0232432.
 PR 14-SEP-2000; 2000US-0232433.
 PR 14-SEP-2000; 2000US-0232434.
 PR 14-SEP-2000; 2000US-0232435.
 PR 14-SEP-2000; 2000US-0232436.
 PR 14-SEP-2000; 2000US-0232437.
 PR 14-SEP-2000; 2000US-0232438.
 PR 14-SEP-2000; 2000US-0232439.
 PR 14-SEP-2000; 2000US-0232440.
 PR 14-SEP-2000; 2000US-0232441.
 PR 14-SEP-2000; 2000US-0232442.
 PR 14-SEP-2000; 2000US-0232443.
 PR 14-SEP-2000; 2000US-0232444.
 PR 14-SEP-2000; 2000US-0232445.
 PR 14-SEP-2000; 2000US-0232446.
 PR 14-SEP-2000; 2000US-0232447.
 PR 14-SEP-2000; 2000US-0232448.
 PR 14-SEP-2000; 2000US-0232449.
 PR 14-SEP-2000; 2000US-0232450.
 PR 14-SEP-2000; 2000US-0232451.
 PR 14-SEP-2000; 2000US-0232452.
 PR 14-SEP-2000; 2000US-0232453.
 PR 14-SEP-2000; 2000US-0232454.
 PR 14-SEP-2000; 2000US-0232455.
 PR 14-SEP-2000; 2000US-0232456.
 PR 14-SEP-2000; 2000US-0232457.
 PR 14-SEP-2000; 2000US-0232458.
 PR 14-SEP-2000; 2000US-0232459.
 PR 14-SEP-2000; 2000US-0232460.
 PR 14-SEP-2000; 2000US-0232461.
 PR 14-SEP-2000; 2000US-0232462.
 PR 14-SEP-2000; 2000US-0232463.
 PR 14-SEP-2000; 2000US-0232464.
 PR 14-SEP-2000; 2000US-0232465.
 PR 14-SEP-2000; 2000US-0232466.
 PR 14-SEP-2000; 2000US-0232467.
 PR 14-SEP-2000; 2000US-0232468.
 PR 14-SEP-2000; 2000US-0232469.
 PR 14-SEP-2000; 2000US-0232470.
 PR 14-SEP-2000; 2000US-0232471.
 PR 14-SEP-2000; 2000US-0232472.
 PR 14-SEP-2000; 2000US-0232473.
 PR 14-SEP-2000; 2000US-0232474.
 PR 14-SEP-2000; 2000US-0232475.
 PR 14-SEP-2000; 2000US-0232476.
 PR 14-SEP-2000; 2000US-0232477.
 PR 14-SEP-2000; 2000US-0232478.
 PR 14-SEP-2000; 2000US-0232479.
 PR 14-SEP-2000; 2000US-0232480.
 PR 14-SEP-2000; 2000US-0232481.
 PR 14-SEP-2000; 2000US-0232482.
 PR 14-SEP-2000; 2000US-0232483.
 PR 14-SEP-2000; 2000US-0232484.
 PR 14-SEP-2000; 2000US-0232485.
 PR 14-SEP-2000; 2000US-0232486.
 PR 14-SEP-2000; 2000US-0232487.
 PR 14-SEP-2000; 2000US-0232488.
 PR 14-SEP-2000; 2000US-0232489.
 PR 14-SEP-2000; 2000US-0232490.
 PR 14-SEP-2000; 2000US-0232491.
 PR 14-SEP-2000; 2000US-0232492.
 PR 14-SEP-2000; 2000US-0232493.
 PR 14-SEP-2000; 2000US-0232494.
 PR 14-SEP-2000; 2000US-0232495.
 PR 14-SEP-2000; 2000US-0232496.
 PR 14-SEP-2000; 2000US-0232497.
 PR 14-SEP-2000; 2000US-0232498.
 PR 14-SEP-2000; 2000US-0232499.
 PR 14-SEP-2000; 2000US-0232500.

PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-541565/60.
 DR N-PSDB; ABA11094.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 11; SEQ ID NO 3425; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 101 AA;
 Query Match 16.8%; Score 523; DB 22; Length 101;
 Best Local Similarity 99.0%; Pred. No. 4.2e-43;
 Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 291 AEQLINPFGEDDDPFTNWIVDRLNQLVSLAVDENHQLPRMEPDMYNNKPPQPYTAA 350
 DB 1 AEQLINPFGEDDDPFTNWIVDRLNQLVSLAVDENHQLPRMEPDMYNNKPPQPYTAA 60
 QY 351 SAQFRASFMGSTFNISLNKEEMEFQPNQDEEDAHAG 388
 DB 61 SAQFRASFMGSTFNISLNKEEMEFQPNQDEEDAHAG 98
 RESULT 15
 ABOG08608
 ID ABOG08608 standard; Protein; 701 AA.
 XX
 AC ABOG08608;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #8599.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

Search completed: July 10, 2003, 12:24:10
Job time : 38.1203 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2003, 12:20:43 ; Search time 9.68173 Seconds
(without alignments)
2506.127 Million cell updates/sec

Title: US-09-622-964-3

Perfect score: 3120

Sequence: 1 MTTTYSQVANRLGFSRL.....TLKHMDPFWALENDEAHS 585

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	99.7	585	1 VMD2 HUMAN	O76090 homo sapien
2	752	24.1	632	1 YQ22 CAEEL	P34672 caenorhabdi
3	751	24.1	405	1 YQ33 CAEEL	O09379 caenorhabdi
4	717	23.0	499	1 YHDI CAEEL	Q23369 caenorhabdi
5	682	21.9	413	1 YV4Q CAEEL	O45435 caenorhabdi
6	679	21.8	450	1 YQ23 CAEEL	Q17528 caenorhabdi
7	669.5	21.5	523	1 YQ84 CAEEL	Q17529 caenorhabdi
8	656	21.0	459	1 YK78 CAEEL	P34319 caenorhabdi
9	646.5	20.7	513	1 YXAK CAEEL	Q21373 caenorhabdi
10	636	20.4	456	1 YN4 CAEEL	P34577 caenorhabdi
11	622.5	20.0	501	1 YSV1 CAEEL	Q22566 caenorhabdi
12	622	19.9	420	1 YCEL CAEEL	Q94175 caenorhabdi
13	585	18.8	400	1 YV6L CAEEL	Q19978 caenorhabdi
14	503.5	16.1	884	1 YAVK CAEEL	Q17851 caenorhabdi
15	474.5	15.2	602	1 YH34 CAEEL	O18303 caenorhabdi
16	441.5	14.2	411	1 YH55 CAEEL	O45363 caenorhabdi
17	395	12.7	434	1 YSWJ CAEEL	O45363 caenorhabdi
18	390	12.5	81	1 VMD2 MOUSE	O88870 mus musculus
19	127.5	4.1	315	1 YNEE-SALTI	Q82706 salmonella
20	127.5	4.1	315	1 YNEE-SALTY	Q82706 salmonella
21	115.5	3.7	613	1 YBID VIBCH	Q9kv48 vibrio chol
22	113	3.6	932	1 YAJA SCHPO	Q09897 schizosacch
23	111.5	3.6	798	1 YBPA HUMAN	Q14694 homo sapien
24	107	3.4	389	1 O85C DROME	Q9vbg6 drosophila
25	107	3.4	961	1 YF02 YEAST	P54860 saccharomyc
26	105	3.4	518	1 ATEL HUMAN	O95260 homo sapien
27	104.5	3.3	457	1 SECT CHLTR	P78539 chlamydia t
28	104.5	3.3	741	1 YBIO ECOLI	P75833 escherichia
29	104.5	3.3	1215	1 BGN DROME	Q9w112 drosophila
30	103	3.3	321	1 YNEE DROME	Q8xaz3 escherichia
31	102.5	3.3	321	1 YNEE ECOLI	P76146 escherichia
32	101.5	3.3	1385	1 CTA1 MOUSE	O54991 mus musculus
33	101	3.2	560	1 EFS_MOUSE	O64355 mus musculus

34	100.5	3.2	577	1 VGP EBORS	O98853 ebola virus
35	100.5	3.2	998	1 CBF_HUMAN	Q03701 homo sapien
36	100	3.2	992	1 EBN6_EBV	P03204 Epstein-barr
37	99.5	3.2	1500	1 SSP5_STRGN	P16952 streptococc
38	99.5	3.2	2195	1 SC16 YEAST	P48415 saccharomyc
39	99	3.2	309	1 Y716_RHILO	Q98866 rhizobium l
40	99	3.2	342	1 VG74_KSHV	Q98146 kapoei's sa
41	98.5	3.2	681	1 CRY1_ARATH	Q43125 arabidopsis
42	98	3.1	307	1 Y787_ANASP	Q8yusd anabaena sp
43	98	3.1	2418	1 SPCA_HUMAN	P02549 homo sapien
44	97.5	3.1	713	1 TUP1 YEAST	P16649 saccharomyc
45	97.5	3.1	1384	1 CTA1_HUMAN	P78357 homo sapien

ALIGNMENTS

RESULT 1
VMD2_HUMAN
ID VMD2_HUMAN STANDARD; PRT; 585 AA.
AC O76090; O75904;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bestrophin (Viteliform macular dystrophy protein) (TUI5B).
GN VMD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS BMD.
RX MEDLINE=98367043; PubMed=9700209;
RA Marquardt A., Stoehr H., Passmore L.A., Kraemer F., Rivera A.,
RA Weber B.H.P.;
RT "Mutations in a novel gene, VMD2, encoding a protein of unknown
RT properties cause juvenile-onset vitelliform macular dystrophy (Best's
RT disease).";
RL Hum. Mol. Genet. 7:1517-1525(1998). *Sept*
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS BMD P-6; H-85; C-93; N-227 AND E-299.
RX MEDLINE=98324772; PubMed=96623195;
RA Petrukhin K., Koisti M.J., Bakall B., Li W., Xie G., Marknell T.,
RA Sandgren O., Forstman K., Holmgren G., Andreasson S., Vujic M.,
RA Bergen A.A.B., McCarty-Dugan V., Figueroa D., Austin C.P.,
RA Metzker M.L., Caskey C.T., Wadelius C.;
RT "Identification of the gene responsible for Best macular dystrophy";
RL Nat. Genet. 19:241-247(1998).
RN [3]
RP VARIANTS BMD HIS-13; CYS-93; CYS-218; ASP-300; GLU-301 AND ILE-307.
RX MEDLINE=99265978; PubMed=10331951;
RA Caldwell G.M., Kakuk L.B., Griesinger I.B., Simpson S.A., Nowak N.J.,
RA Small K.W., Maunee I.H., Rosenfeld P.J., Sieving P.A., Shows T.B.,
RA Ayagari R.;
RT "Bestrophin gene mutations in patients with Best vitelliform macular
RT dystrophy";
RL Genomics 58:98-101(1999).
RN [4]
RP VARIANTS BMD V-10; V-82; C-92; H-96; S-135; C-218; S-218 AND K-293.
RX MEDLINE=99320852; PubMed=10394929;
RA Bakall B., Marknell T., Ingvas S., Koisti M.J., Sandgren O., Li W.,
RA Bergen A.A.B., Andreasson S., Rosenborg T., Petrukhin K., Wadelius C.;
RT "The mutation spectrum of the bestrophin protein -- functional
RT implications";
RL Hum. Genet. 104:383-389(1999).
RN [5]
RP VARIANTS BMD/BMD, VARIANT AVMD K-146, AND VARIANT BULL'S EYE Q-119.
RX MEDLINE=99381534; PubMed=10453731;
RA Allikmets R., Seddon J.M., Bernstein P.S., Hutchinson A., Atkinson A.,
RA Sharma S., Gerrard B., Li W., Metzker M.L., Wadelius C., Caskey C.T.,
RA Dean M., Petrukhin K.;
RT "Evaluation of the Best disease gene in patients with age-related
RT macular degeneration and other maculopathies.";

Query Match 99.7%; Score 3110; DB 1; Length 585;
 Best Local Similarity 99.7%; Pred. No. 4.7e-225;
 Matches 583; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLLCYIIIRFYRLALTEEQOL 60
 DB 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLLCYIIIRFYRLALTEEQOL 60
 QY 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEKDEQ 120
 QY 121 SLLRLRTLIIRYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 DB 121 GRLLRLRTLIIRYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 QY 181 FWVPWVFWANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCGLHYAYDWISIPLVYTVQV 240
 DB 181 FWVPWVFWANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCGLHYAYDWISIPLVYTVQV 240
 QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 QY 301 DDDDPETNWIIDRNLOVSLLADEMHQDLPRMEPDWYKPKPPPYTAASQFRASPM 360
 DB 301 DDDDPETNWIIDRNLOVSLLADEMHQDLPRMEPDWYKPKPPPYTAASQFRASPM 360
 QY 361 GSTFNISLNK--EEMEFQPNQDEDEAHAG-----IIGRFLGLQSHDHHPPRANR 420
 DB 361 GSTFNISLNK--EEMEFQPNQDEDEAHAG-----IIGRFLGLQSHDHHPPRANR 420
 QY 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPTEMFPP 480
 DB 421 LHEGLPKNHKAQKQNVRGQEDNKAWLKAVDAFKSGPLYQRPYYSAPOTPLSPTEMFPP 480
 QY 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKSKFELLSSDGLMBHEPVSQVRRKTVFN 540
 DB 481 LEPAPSCLKHVTGIDTKDKSLKTVSSGAKSKFELLSSDGLMBHEPVSQVRRKTVFN 540
 QY 541 LTDMPETNHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585
 DB 541 LTDMPETNHLKPELQSPNTNIHTTLKHMDPYWALENRDEAHS 585

RESULT 2
 Y022 CAEEL STANDARD; PRT; 632 AA.
 AC P34672;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK688.2 in chromosome III.
 GN ZK688.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins J., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showken R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,

Wohlman P.;
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 CC 1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 EMBL; LI6621; AAA28228.1; -;
 DR PIR; S44917; S44917.
 DR WormPep; ZK688.2; CE00460.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 73829 MW; 0ABDD1755EF11642 CRC64;
 Query Match 24.1%; Score 752; DB 1; Length 632;
 Best Local Similarity 35.9%; Pred. No. 1.2e-48;
 Matches 152; Conservative 94; Mismatches 163; Indels 14; Gaps 4;

QY 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLLCYIIIRFYRLALTEEQOL 60
 DB 1 MTITTSQVANARLGSFSLLCWGRSIIKLYGEEFLIPLLCYIIIRFYRLALTEEQOL 60
 QY 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEKDEQ 120
 DB 61 MPEKLTLYCDSYIOLIPISFVLGYVTLVTRWNOYENLWPDRLMSLVSGFVEKDEQ 120
 QY 121 SLLRLRTLIIRYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 DB 121 SLLRLRTLIIRYANLGNVILRSVSTAVYKRFPSAQHLVQAGWTPPAHKOLEKLSLPHNM 180
 QY 181 FWVPWVFWANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCGLHYAYDWISIPLVYTVQV 240
 DB 181 FWVPWVFWANLSMKAWLGGRIIDPILLOSLNEMNTLRTQCGLHYAYDWISIPLVYTVQV 240
 QY 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 DB 241 TVAVYSFELTCLVGRQFLNPAKAYPGHELDLVVPVFTLQFFYYGWLKVAEQLINPFG 300
 QY 301 DDDDPETNWIIDRNLOVSLLADEMHQDLPRMEPDWYKPKPPPYTAASQFRASPM 360
 DB 301 DDDDPETNWIIDRNLOVSLLADEMHQDLPRMEPDWYKPKPPPYTAASQFRASPM 360
 QY 361 GSTFNISLNK--EEMEFQPNQDEDEAHAG-----IIGRFLGLQSHDHHPPRANR 409
 DB 361 GSTFNISLNK--EEMEFQPNQDEDEAHAG-----IIGRFLGLQSHDHHPPRANR 409
 QY 410 TKL 412
 DB 418 TSL 420

RESULT 3
 YS63 CAEEL STANDARD; PRT; 405 AA.
 ID YS63 CAEEL
 AC Q09379;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ZK675.3 in chromosome II.
 GN ZK675.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 ON NCBI_TaxID=6239;

```

RN  SEQUENCE FROM N.A.
RC  STRAIN=Bristol N2;
RA  Sims M.;
RL  Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z68753; CAA92989.1; -
DR  WormPep; ZK675.3; CE01722.1; -
DR  InterPro; IPR000615; Worm_fam_8.
DR  Pfam; PF01062; DUF289; 1.
DR  ProDom; PD002802; Worm_fam_8; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 405 AA; 47835 MW; 0F589D874E9E0561 CRC64;

Query Match      24.1%; Score 751; DB 1; Length 405;
Best Local Similarity 39.2%; Pred. No. 7.6e-49;
Matches 153; Conservative 79; Mismatches 140; Indels 18; Gaps 5;

QY  1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB  1 MTISYSD-----TFLLKLLFRWKSGLWKAIWKHLILFTWYIINAYYRFGMTKEQQN 52

QY  61 MPEKLTLYCDSYIQI-IPISFVLGYVTLVVTRWNOYENLPDRMLSLVSGFVEGKQ 120
DB  53 EPIKYVMDVGTWKRIPLTFLLGFGYAMIVRWDCCLISWFDHLLVNVSAIRGQDPE 112

QY  121 SLLRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSL- - -P 177
DB  113 TRIIRKTIARYILTSVLAWRSISRLVARYPTDHLVDSGLMTKEEMVFKSILVHVP 172

QY  178 HMFVWVWVFWANLMSKAWLGGRIKDPILLOSLANEMNTLRQCCHLYAYDNISPLVYTV 237
DB  173 HOKWVPLNWIQTMVRCPEKGLTHITNELRVLLDALEKRYGFFOLFIDYIAIPLVYT 232

QY  238 QVTVVAVYFFLTCVGRFLNPAKAYPGHELDLVVPVFTFLQFFPYGWLKVAQOLNP 297
DB  233 QVSTISVGYFLALIGRY--PSKNENEIVDVVYPIFTILOFLFYVGLVAGEDLMFP 290

QY  298 FGEDDDDFETNWIVDRLNQLVSLAVDEMHDQLPRMEPDMYWNKPFPQPPYTAASQPRRA 357
DB  291 FGADEDFEFNYILERNLEVSMLIVDELHNQVPPVYVESL--DDEIRLLHTSASSKLSNH 348

QY  358 SPMGSTFNISLNKEMEFQP--NQDEED 384
DB  349 FORQHLRLKLFNVDAKQVAVPCNHLNERD 378

RESULT 4
YHDI CAEEL STANDARD; PRT; 499 AA.
AC Q23359;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZC518.1 in chromosome IV.
GN ZC518.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC STRAIN=Bristol N2;
RA Thomas K.;

Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
-!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; Z68753; CAA92989.1; -
WormPep; ZC518.1; CB06601.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; DUF289; 1.
ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 499 AA; 57229 MW; 4E87725437A5C9DF CRC64;

Query Match      23.0%; Score 717; DB 1; Length 499;
Best Local Similarity 35.2%; Pred. No. 3.5e-46;
Matches 160; Conservative 78; Mismatches 158; Indels 58; Gaps 9;

QY  1 MTITYTSQVANARLGSFRLLCWRGSIYKLLYGEFLIFLLCYIIRFYRLALTEEQQL 60
DB  1 MTISYTLDSQTNLQSPFSLLRWGSVWKAIVFGQLAVTAVFLILSCIYRYMLSPSQD 60

QY  61 MPEKLTLYCDSYIQI-IPISFVLGYVTLVVTRWNOYENLPDRMLSLVSGFVEGKDE 119
DB  61 VFEQLIRYFDNKLDAIPLTFLLGFGFVSFVWARGSIILNGIGWIDASLLFATYIRGAD 120

QY  120 OSRLRLTIRYANLGNVILRSVSTAVYKRPSPSAQHLVQAGFMTPAEHKQLEKLSLPHN 179
DB  121 ETRVTRRLRVLYLSQALVLRDISMQRKRPPTMTLAASGLMTHEDMILIDHTKDPYS 180

QY  180 MFWVWVWVFWANLMSKAWLGGRIKDPILLOSLANEMNTLRQCCHLYAYDNISPLVYTV 239
DB  181 RYWTISQWLSNLVYECQKGVDSYLLNKNIVDEIGKFRHGLASLLKYDWPVPLVYQV 240

QY  240 VTVAVYFFLTCVGRQFL-----NPAKAYPGHELDLVVPVFTFLQFFPYGWLKVAEQI 295
DB  241 IFLAVRIYFMICLIGRQFIVTGPNS-----GIDLWLPITTWQFLVYVGMGMKVAEALL 294

QY  296 NPFGEDDDDFTNWIVDRLNQLVSLAVDEM--HQD-----LPRMEPDMYWNKP 341
DB  295 NPLGEDDDDLSCVYIIDKNLITGLSIVDTMKHDDDTGYSMVEEHNAKTPAQKQDEFGMD 354

QY  342 EPQPPYTAASQPRRAFMGSTFNISL--NKEEMEFQP-----NQ 379
DB  355 KIAPLYSMESAERSVHPLVGSASKINLVKNKEIVMTPHKNKLSLSDPSEQTYLRRVNV 414

QY  380 EDEDAHAGIIGRFLGLOSHDHPHPRANSRTKLL 413
DB  415 SDHNAKGA---KQGLS-----RANSPDKCL 437

RESULT 5
YV4Q CAEEL STANDARD; PRT; 413 AA.
AC O45435;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein F32B6.9 in chromosome IV.
GN F32B6.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Basham V.;
```

```

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z81074; CAB03043.1; --
DR WormPep; F32B6.9; CE09864.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 413 AA; 48965 MW; A6E69A83C78790B CRC64;
-----
Query Match 21.9%; Score 682; DB 1; Length 413;
Best Local Similarity 39.6%; Pred. No. 1.1e-43;
Matches 141; Conservative 67; Mismatches 126; Indels 22; Gaps 7;
QY 1 MTITTSQVANARLGSFRLLCWGSIVKLYGFLIFLLCYIIRFYRLAL-----T 55
DB 1 MTISYS-----GNVIRLLRWKGSIVRTAKELLIIILYISVRVYLLKGLDIDDD 52
QY 56 EQQL-----MFEKLTLYCDSYIQLIPISVLGFYVTVLVVTRWNOYENLPWDRMLSLVS 111
DB 53 EDRRLKMRMEFTFCQDCDSYTRLLPLTLFLGFGYSNVVWVRWQFETLYWPELILSVLC 112
QY 112 GFVSGKDFQSLRLRLTRLYANLGNVILRSVSTAVYKRPSPQAHLVQAGFWTPAEHKQL 171
DB 113 TVLHQHDEKSKRRRTIARYLNLALAWRDISSKIRLRFPSVHLSGLLTKKEYQIL 172
QY 172 EKLSLPH-NMPWVPWVWFANLSMKAWLGGRIIDPILLOSLLNEMNTLRTQCGHLVAYDWI 230
DB 173 EAMAEVSSSWITPLHWIQLIMRQVEEHKPTASLNFQVGLRI FROSLRKLVSYDMV 232
QY 231 SIPLVYQVTVVAVYSFFLTCLVGRQFLNPAKAYFGHELDLVVPVFTFLQFFFYVGLVKV 290
DB 233 CVPLVYQVVAALATYSPFFFTLFGQPLFP-DIETGKELDLVVPVFTVQLFFVGVFKV 291
QY 291 AEQLINFGEDDDDFETNWIIVDRNLQVSLAVDEMHDQD-LPRME--PDMYNKPKSP 343
DB 292 QGLMRPFGLDLDDIELNYILDRNVRISFAIVNQLQESPIPDFESNDDKLWHEHP 347
-----
RESULT 6
YOE3 CAEEL STANDARD; PRT; 450 AA.
AC Q17528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.3 in chromosome IV.
GN B0564.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73422; CAA97765.1; --
DR WormPep; B0564.3; CE05177.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 450 AA; 53275 MW; A0PED9A476166AD7 CRC64;
-----
Query Match 21.8%; Score 679; DB 1; Length 450;
Best Local Similarity 32.9%; Pred. No. 2.1e-43;
Matches 155; Conservative 80; Mismatches 184; Indels 52; Gaps 9;
QY 1 MTITTSQVANARLGSFRLLCWGSIVKLYGFLIFLLCYIIRFYRLALTEEQQL 60
DB 1 MTINHKETWTSHPWTFFLLFKWGSIVKAVYMETIIFLICYIISVIYKTAMGESSQR 60
QY 61 MFEKLTLYCDSYIQLIPISVLGFYVTVLVVTRWNOYENLPWDRMLSLVSFGFVEGKDEQ 120
DB 61 TFEGLRVYFDKLSVIPLEFVLGFVTVVNRWTKLYQTIGFDNVGLMANCYIRGATEK 120
QY 121 SLLRLTRLYANLGNVILRSVSTAVYKRPSPQAHLVQAGFWTPAEHKOLEKLSLPHN- 179
DB 121 ARIYERNIMRYCELQVILVFRDMSWETRRFRFTMETVVAAGFPMKHELELYNSYDTKYNS 180
QY 180 ----MFWVWVWFANLSMKAWLGGRIIDPILLOSLLNEMNTLRTQCGHLVAYDWISPLV 235
DB 181 KLGTWVTPANWALCMTYKARKDGYIESDYFKAQMEGEIRTWRTNIETWVCNVDWVPLPM 240
QY 236 YQVTVVAVYSFFLTCLVGRQFLNPAKAYFGHELDLVVPVFTFLQFFFYVGLVKVABOLI 295
DB 241 YPOLVCLANVLYSVIIARQ-LVIEKHQWDEVQVYFPVMTFLQIFIFYGWLKVIDVML 299
QY 296 NPFGEEDDDDFETNWIIVDRNLQVSLAVDEMHDQD-LPRMEPDMYNKPEPPPYTAASAQFR 355
DB 300 NPFGEEDDDDFETNALIDRNITMGLIADN-PMSTPELRKDPFYDEVDPVLLYSESSNIP 358
QY 356 RASFGMSTFNISLN-----KSEMEFQPNQDEDEDAHAGIIGFLGL 396
DB 359 NHHYHGSVSEVLEQKGNAPVMMPHSGSAANLRMMSPKSVDEDEKONAFSM----- 412
QY 397 QSHDHPHPRANSRTKLLMPKRESLLHGLPKNHKAQKQNVRGQEDNKAWKL 447
DB 413 -SHDD-ARRN---W--REVSLDSSF-----LADLNENKWKI 443
-----
RESULT 7
YOE4 CAEEL STANDARD; PRT; 523 AA.
AC Q17529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein B0564.4 in chromosome IV.
GN B0564.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lightning J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: Z73422; CAA97766.1; -
 CC WormPep; B0564.4; CE05178.
 CC InterPro; IPR000615; Worm_fam_8.
 CC Pfam; PF01062; DUF289; 1.
 CC ProDom; PD002802; Worm_fam_8; 1.
 CC KW Hypothetical protein.
 CC SEQUENCE 53 AA; 61725 MW; EC37F07253E5F9A6 CRC64;

Query Match 21.5%; Score 669.5; DB 1; Length 523;
 Best Local Similarity 36.5%; Pred. No. 1.3e-42;
 Matches 142; Conservative 73; Mismatches 165; Indels 9; Gaps 4;
 QY 1 MTITYTSOVANARLGSRLLLCWRGSIYKLLYGFLPLCYIIRFYRLALTEEQ 60
 DB 1 MTINHKETKSHTWKFFVLLFRWKGSIWKALYMETIIFLYIGLISVYRTAMSEPSQR 60
 QY 61 MFEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFVGEKDEQ 120
 DB 61 TFSVIRYCDKRLSFIPLFVLPVTVDRWTKLWRTVGFIDVCLLANLYVGTSEK 120
 QY 121 SRLRLRLIRVANGLVILRSVAVYKRPSPSAQHLVQAGFMTPEAHKQLEKLSLPHN- 179
 DB 121 AIIYRRNIARYCALTLVFRDVSMTTRRPPTMETVVAAGFMSKDELGLYNSYTTKNS 180
 QY 180 ---MFVVPWVWFWANLGMKALGRIKIDPILLOSLLNEMNTLRTQCGHLYAYDMISIP 235
 DB 181 RLKXKXIPANWALCMYKARKDGYSDYFKAQMEGBIRTWRTNIEWCNVDWVPLPLM 240
 QY 236 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQ 295
 DB 241 YPQLVCLAVNLYFLVSIARQ-LVIEKHWDVDEVDPVFTFLQFFYVGVMLKVAEQ 299
 QY 296 NPFGEDEDDFTNNIVDRNQLVSLAVDEMHDLPMEPDPMYKPEQPPYTAASQPR 355
 DB 300 NPFGEDEDDFTNALIDRNITMGLKWDNT-MKTELLKQDFDREVLVSLLYSESSQIS 358
 QY 356 RASPMGTFNLSLAKKE--EMEFQPNQDE 382
 DB 359 NYHVGSTSEVHLQKSSVRMIPHSQSE 387

RESULT 8
 YKX8 CAEEL STANDARD; PRT; 459 AA.
 AC P34319;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C07A9.8 in chromosome III.
 GN C07A9.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulten L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Sime M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sultston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans.";
 RL Nature 368:32-38 (1994).
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

EMBL: Z29094; CAA82342.1; -
 DR PIR; S40708; S40708.
 DR WormPep; C07A9.8; CE00497.
 DR InterPro; IPR000615; Worm_fam_8.
 DR Pfam; PF01062; DUF289; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 459 AA; 53570 MW; C2D8F69E4078BF9F CRC64;

Query Match 21.0%; Score 656; DB 1; Length 459;
 Best Local Similarity 32.3%; Pred. No. 1.1e-41;
 Matches 129; Conservative 91; Mismatches 151; Indels 28; Gaps 5;
 QY 1 MTITYTSOVANARLGSRLLLCWRGSIYKLLYGFLPLCYIIRFYRLALTEEQ- 59
 DB 44 LSYNNYDLATSKSLMIVRMIFKWRGSLVQAVYKELIVMICAYSLSVIYRFPALTRSOKE 103
 QY 60 -----LMPEKLTLYCDSYIQLIPISFVLGFTVTLVTRWNNQYENLPWDRMLSLVSGFV 114
 DB 104 QNKEIIIFEREFGYCDARMGYLPLNFVLGFFCNIIIRRLKLYTSLGNIDNIALFVSAY 163
 QY 115 EGKDBOSLLRLRLIRVANGLVILRSVAVYKRPSPSAQHLVQAGFMTPEAHKQLEK 174
 DB 164 RTDDRAQIRNIIRYICVISOCLVFRDTHVGVRRRFFPLEAVQAGIMLPHELEKFSI 223
 QY 175 SLPHNMFVVPWFWANLGMKALGRIKIDPILLOSLLNEMNTLRTQCGHLYAYDMISIP 234
 DB 224 KSRQYKXVSWFNWALELLNVAKTEKSIDGNARNAIAQEISKFRSALTIVSWYDVP 283
 QY 235 YTVQVTVAVYSFFLTCLVGRQFLNPAKAYPGHELDLVVPVFTFLQFFYVGVMLKVAEQ 294
 DB 284 MYPQLVNNVAHTYFPLCTFTRQFFISADAHNKTEVDLYIPFTIIEFIYMGWLKVAMEL 343
 QY 295 INPGEDEDDFTNNIVDRNQLVSLAVDEMHDLPMEPDPMYKPEQPP- 346
 DB 344 LNPFGEDADDPCNLLIDRNLAIGLTSVDDAYDQLPEVKPDVFTGSGVKPLSDSDTRSLK 403
 QY 347 -YTAASQFRASPMGTFNLSLAKKEEMEF-----QPNQ 379
 DB 404 YHFGSAAQMEISY-----LKKEENKMAAGKKPK 434

RESULT 9
 YXAK CAEEL STANDARD; PRT; 513 AA.
 AC Q21973;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein R13.3 in chromosome IV.
 GN R13.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Matthews P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.


```
RESULT 13
YV6L CAEEL STANDARD; PRT; 400 AA.
ID YV6L CAEEL
AC Q19978;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein F3268.4 in chromosome V.
GN F3268.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 272509; CAA96648.2; -.
DR WormPep; F3268.4; CE23700.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 400 AA; 46750 MW; B1E8ABEC3B862E84 CRC64;
Query Match 18.8%; Score 585; DB 1; Length 400;
Best Local Similarity 31.5%; Pred. No. 1.9e-36;
Matches 140; Conservative 83; Mismatches 175; Indels 46; Gaps 11;
QY 1 MTITYTSQVANARLGSFRLLLCMRGSYKLYGFLFLCYIIRFYRLALTEEQOL 60
DB 1 MTISYDEE-----FSLMLRWGSIWKAVLKDGLIGFYIAYIVLAFQVLLDEKGE 52
QY 61 MFEKLTLYCDSYIQLIPISFVLGYVTVLVTWRMNQYENLPWDRMLSLVSGFVEQDEQ 120
DB 53 YFTGWIMWCEIGAQYIPLSFLGFPVSLIVARWWEQNCISWPDQWIMVSAQLPG--NE 110
QY 121 SRLARRTLIRYANLGNVILRSVSTAVYKRPFAQHLVQAGFMTPAEHKQLEKLSLPHN 180
DB 111 MNVVRQTIARWSSQAAIAWAGSVSKTLKRPFTERRHVMASKLMTBEEYDLYNMTDAPHGK 170
QY 181 FWVFWVWFWANLGMKAWLGGRIKDPILLOSLLNEMNTLRQCHLYAYDWISIPLVYTVQV 240
DB 171 WFIILMTVNLVKKQKQG--IIDSQMDMLLKQVYSYRDGFAMLFYVDWIKIPLVYTVQV 229
QY 241 TVAVYSFFLTCVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGMVGLKVAEQLINPFE 300
DB 230 AIATYGVFFICLIGKQPKLDQRSME-KEITILPFIPTFFQMLFYGLWLVKVGQ-----PSI 283
QY 301 DDDDFETNIVDRNLQVSLAVDEKHQDLPRMEDMTYKNKEPOPPYTAASQAQFRASFM 360
DB 284 RKKTSELNLYLDRNTAHTAHMASLSQDLPISGAPV-----PAVPHTRASFKIQDVIPK 338
QY 361 GSTFNISLNKEMBE-FQFNQDEEDAHAGIIGRFLGQSHDHPHPRANSRTKLLWPKRES 419
DB 339 SHLAGPKLSEAMKLIKP-----ED-----LEEHE-----RLMEETKVTNRORLQ 378
QY 420 LHEGLPKQNHKAAKQNVRGQEDNK 443
DB 379 TLVRAL---EKSRTNATINEDDE 399
RESULT 14
YAVK CAEEL STANDARD; PRT; 884 AA.
ID YAVK CAEEL
AC Q17851;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C09B9.3 in chromosome IV.
GN C09B9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bradshaw H.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U50069; AAB37559.2; -.
DR WormPep; C09B9.3; CE29571.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 2.
DR ProDom; PD002802; Worm_fam_8; 3.
DR KW Hypothetical protein; Transmembrane; Repeat.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
SQ SEQUENCE 884 AA; 102851 MW; 8DB83F9699B8FAID CRC64;
Query Match 16.1%; Score 503.5; DB 1; Length 884;
Best Local Similarity 32.7%; Pred. No. 7.1e-30;
Matches 105; Conservative 70; Mismatches 139; Indels 7; Gaps 4;
QY 1 MTITYTSQVANARLGSFRLLLCMRGSYKLYGFLFLCYIIRFYRLALTEEQOL 60
DB 1 MTISYTVDAVESYFGFPFKVLPFKWGSVKLIHRELFMWLVLYTVLAIYR-TLDEERK 59
QY 61 MFEKLTLYCDSYIQLIP--ISFVLGYVTVLVTWRMNQYENLPWDRMLSLVSGFVEGKD 118
DB 60 IFRS---NIEHFINFEPISILTFMLSFFVTIVQWRNNVFTNMGFTIENAAAYASVSEPMK-NG 115
QY 119 EQSILLRRTLRYANLGNVILRSVSTAVYKRPFAQHLVQAGFMTPAEHKQLEKLSLPH 178
DB 116 EDVRAQRTVIRYLVASQILVWRSISIKALRRFNYESIVTAGFTKGEESTIQTNDLSY 175
QY 179 NMFWVWVWFWANLGMKAWLGGRIKDPILLOSLLNEMNTLRQCHLYAYDWISIPLVYTVQ 238
DB 176 DSSCVPIRWAIOQLRHQVRSNGNFFSHSVYRATWKEVSDFEIHLKRVKVDWVPIPLAYPQ 235
QY 239 VTVAVYSFFLTCVGRQFLNPAKAYPGHELDLVVPVFTFLOFFFYVGMVGLKVAEQLINP 298
DB 236 VIFFAVRLYFVICAPAKQYFDLDDDDARYVIHYHYPFIVTVFQFICLMGWLKVAEALLNPL 295
```



```
QY 299 GEDDDDFETNWIIVDRNLQVSL 319
Db 296 GEDDDDFEVNFLIDSNIYTM 316

RESULT 15
YHS4_CAEEL STANDARD; PRT; 602 AA.
AC O18303;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK849.4 in chromosome I.
GN ZK849.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
REVIEWS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82095; CAB05027.2; -
DR WormPep; ZK849.4; CE25696.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 2.
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 67607 MW; 458AB78802BD6B3A CRC64;

Query Match 15.2%; Score 474.5; DB 1; Length 602;
Best Local Similarity 29.5%; Pred. No. 6.2e-28;
Matches 135; Conservative 65; Mismatches 147; Indels 111; Gaps 12;

QY 1 MTITVTSQVANARLGSRLLCWRGSIYKLYGBFLIFLCYIIRFYRLAL-----T 55
Db 14 MVTYNNRAVSTESIHNFVSIWYHSGSLVKSISKEYAIWLVLYHLFIYRVVMPFGWA 73
QY 56 BEQQLMFEKLTLYCDSYLQLPISVLGIVYTLVVTWNNQYENLPWDRMLSLVSGFVE 115
Db 74 DYCKKVIENWTPHQD---MTIPLEFLGFFVTVIDRWKAFQNPITYIEICAFVAAAIP 130
QY 116 G-----KDEQSR----- 122
Db 131 GRIRIINKQNEVRAPGQDLRDVPAVEVNHVPQVGPAPAAALPMGAPERPAVALPMP 190
QY 123 -----LLRRTLIRYANLGNVLILRSVSTAVYKRPFSQAHLVQAGFMT 164
Db 191 GAPEGPAAIQGPSVDVKLTARRTIIRYLVLSQLILLFREISTVKKRFVDLCLVDSKFLT 250
QY 165 PAEHKOLEK-LSLPH-NMFVWPWW-FANLS-----MKWLGRIIDPILQLSL 211
Db 251 DEELKILSKVKCHYDSYIFINWAPSILOEHKVKSNPEFINAW--NVIRDWQVKLSL 308
QY 212 NEMNTLRQCGLXAYDWISIPLVYTVVTVAVYSFFLTCLVGQFLNPAKAYPGCHELDL 271
Db 309 RN-----GDFIPIPLAYQAVFLAIRFYFLVCLFTROHLDDDK---KSIDY 352
```

```
QY 272 VVPVFTFLQFFYVGLKVAEQLINPFGEDDDDDFETNWIIVDRNLQVSLIIVDEMHDLP 331
Db 353 YFPLMTSLQFIFVGMKVAEILLNPMGEDDDDDFELNNIIDKLYIGLAIVTECGKHP 412
QY 332 MEPMYNNKPSPPPYTAASAQFRASPMWGSTFNISLN 369
Db 413 IVKDTIGKDCLPFFYPQNDNERNRA-LVGSTKNINLS 449
```

Search completed: July 10, 2003, 12:24:43
Job time : 12.6817 secs